ABSTRACT

Title of dissertation: APPLICATION OF ULTRAHIGH RESOLUTION MASS SPECTROMETRY AND DEUTERIUM LABELING TO DETERMINE THE CONTRIBUTION OF KETONE/ ALDEHYDE-CONTAINING SPECIES TO THE COMPOSITION AND OPTICAL PROPERTIES OF DISSOLVED ORGANIC MATTER

Marla Bianca
Doctor of Philosophy, 2020

Dissertation directed by: Professor Neil Blough
Department of Chemistry and Biochemistry

Dissolved organic matter (DOM) is a complex, heterogeneous mixture comprised of thousands of chemical species, found in almost all aquatic environments and is one of Earth's largest carbon reservoirs. DOM is known to affect many biogeochemical processes and may be a crucial component of the global carbon cycle. However, due to the inherent complexity of DOM, understanding and relating its photophysical and photochemical properties to its composition is difficult.

This dissertation describes and applies a mass labeling method using ultrahigh resolution mass spectrometry and deuterium labeling to determine the contribution of ketone/ aldehyde-containing species to the composition and optical properties of DOM. Sodium borodeuteride ($NaBD_4$) selectively and irreversibly reduces ketone/aldehyde groups and the changes due to reduction were observed through mass

spectrometry and ultraviolet/visible (UV-Vis) and fluorescence spectroscopy. The reduction of ketone/aldehyde-containing species by $NaBD_4$ results in loss of absorption and creates unique mass markers (deuterated species at mass $M+3.021927n$) in the mass spectrum. The internal consistency of this algorithm for identifying reduced species was tested using two additional methods, both of which resulted in consistently identified reduced species in the mass spectra.

This method was applied to Suwannee River fulvic acid (SRFA) with increasing concentrations of $NaBD_4$ to evaluate the extent of reduction which differed depending on the concentration used. Additionally, the extent of reduction resulting from the increasing concentrations of $NaBD_4$, was shown to correlate well with changes in the absorption and emission spectra of the corresponding untreated and reduced samples; thus, providing evidence that ketone/aldehyde functional groups contribute substantially to the bulk optical properties of SRFA.

Furthermore, SRFA that was irradiated as well as reduced revealed insights into the structural components in SRFA that were lost or decreased due to irradiation and contributed to the observed optical properties. Also, it was demonstrated that irradiation made terrestrial material appear more like marine DOM.

DOM samples from several different aquatic environments were compared using this method to reveal differences and similarities within the composition of DOM. This method has proven to be a useful tool in relating the changes in the optical properties upon the reduction by $NaBD_4$ to the changes observed by mass spectrometry to reveal information on the composition and well as source and structure of DOM.

APPLICATION OF ULTRAHIGH RESOLUTION MASS SPECTROMETRY AND DEUTERIUM LABELING TO DETERMINE THE CONTRIBUTION OF KETONE/ ALDEHYDE-CONTAINING SPECIES TO THE COMPOSITION AND OPTICAL PROPERTIES OF DISSOLVED ORGANIC MATTER

by

Marla Bianca

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2020

Advisory Committee:
Professor Neil Blough, Chair
Professor Alice Mignerey
Professor Peter Nemes
Professor Michael Gonsior
Professor Akua Asa-Awuku (Dean's Representative)

Acknowledgments

Many people have supported me and given me guidance throughout my graduate career and I am grateful to all of them. I would like to thank Dr. Neil Blough for mentoring and advising me throughout my Ph.D research. Thank you for the experiences and opportunities, such as being on a research vessel and attending conferences, that have helped me grow as a researcher! I would also like to thank Dr. Rossana Del Vecchio, who was a great mentor and for her knowledgeable insights on how to conduct research and skills to better interpret the data.

Thank you to my dissertation committee: Dr. Alice Mignerey, Dr. Peter Nemes, Dr. Michael Gonsior, and Dr. Akua Asa-Awuku. Additional thank you to Dr. Michael Gonsior for helpful guidance, collaboration, and assistance throughout my graduate career. I would also thank Schmitt-Kopplin and Dr. Mourad Harir at the Helmhotz Center for Environmental Sciences.

Thank you to my past and present group members: Dr. Carmen Cartisano, Dr. Daniel Baluha, Danielle Le Roux, Shannon McDonnell, Alejandro Beltran, and Rachel Ashmore. Thank you for being there for insightful discussions as well as fun times and sometimes not so fun times. I am glad I could count on you.

Thank you to Natalia White. I have learned so much from you as a person as well as from being your TA. Thank you for always listening and for your advice over the years, I am so grateful.

I would also like to thank the professors at Moravian College who helped me prepare for graduate school, especially Dr. Diane Husic. I won't forget the all the

hours you helped me with my undergraduate research and your great appreciation for the subject that left a lasting impression on me.

Finally, I thank my family and friends for all their love, support, and patience that made getting through this possible!

Table of Contents

Acknowledgements ii

List of Tables viii

List of Figures ix

List of Abbreviations xiii

1 Aquatic Natural Organic Matter (NOM): Background and Environmental Importance 1
- 1.1 Aquatic Natural Organic Matter (NOM) and Sub-Fractions 1
- 1.2 DOM and CDOM Significance in the Aquatic Environment 5
- 1.3 Optical Properties of CDOM 7
  - 1.3.1 CDOM absorption and emission 7
  - 1.3.2 Quantum Yields 12
  - 1.3.3 Structural Basis of the Optical Properties of CDOM 13

2 Analytical Techniques Used to Study and Characterize DOM: Focus on Ultrahigh Mass Spectrometry 18
- 2.1 Early studies: Structure and Composition of DOM 18
- 2.2 Analytical Techniques for Analyzing DOM 19
- 2.3 Instrumentation for Mass Spectrometric Analysis of DOM 20
  - 2.3.1 Electrospray Ionization of DOM 20
  - 2.3.2 Mass analyzers and ultrahigh resolution 22
- 2.4 Mass spectral Analysis, Interpretation, and Representation of DOM . 25
  - 2.4.1 Mass Distribution 25
  - 2.4.2 Molecular Formula Assignment 26
  - 2.4.3 Presentation of Mass Spectrometric Data 30

3 Mass Labeling Method for Identifying Reduced Peaks and Molecular Formula Assignments (MATLAB Algorithms) 34
- 3.1 Reduction with Sodium Borodeuteride 34
- 3.2 Mass Spectrometric Data Acquisition and Pre-processing 36

3.3 Molecular Formula Assignments and Identification of Reducible Species using MATLAB Algorithms. 37

4 Examining the contribution of ketone- and aldehyde-containing species to the composition and optical properties of Suwannee River fulvic acid as revealed by revealed by ultrahigh resolution mass spectrometry and deuterium labeling 43
4.1 Introduction 43
4.2 Materials and Methods 44
4.2.1 Reagents and Materials 44
4.2.2 Sample Preparation of Suwannee River Fulvic Acid 44
4.2.3 Mass Spectrometric Data Acquisition, Pre-processing, Molecular Formula Assignments, and Identification of Reducible Species in SRFA 46
4.2.4 Optical Measurements for SRFA 46
4.3 Results and Discussion 47
4.3.1 General Mass Spectrometric Features 47
4.3.2 Assessment of Molecular Formulae Algorithm Assignments . . 49
4.3.3 Reduction and Identification of Ketone/aldehyde-containing Species in SRFA. 51
4.3.4 Molecular Composition of Ketone/aldehyde-containing Species in SRFA 57
4.3.5 Relation of Identified Ketone/aldehyde-containing Species in SRFA to the Observed Optical Properties 61
4.4 Conclusions 65

5 Effects of Irradiation and Combined Irradiation *and* Reduction on the Optical and Mass Spectral Properties of SRFA 66
5.1 Introduction 66
5.2 Materials and Methods 68
5.2.1 Reagents and Materials 68
5.2.2 Sample Preparation of SRFA Solution for Irradiation and Reduction 69
5.2.3 Optical Measurements for SRFA: Irradiation and Reduction . 70
5.2.4 Mass spectrometric data acquisition, pre-processing, molecular formula assignments, and identification of reducible species in SRFA and irradiated SRFA 71
5.3 Results and Discussion 72
5.3.1 Effect of Irradiation on the Optical Properties 72
5.3.2 Effect of Irradiation on the Mass Spectrometric Features . . . 74
5.3.3 Effect of Reduction and Combined Irradiation *and* Reduction of SRFA on the Optical Properties 84
5.3.4 Effect of Reduction and Combined Irradiation *and* Reduction of SRFA on the Mass spectrometric Features 85

5.3.5 Relation of Identified Ketone/aldehyde-containing Species in Reduced and Combined Irradiated *and* Reduced SRFA to the Observed Optical Properties After Reduction 90
5.4 Conclusions 92

6 Comparison of the Contribution of Ketone- and Aldehyde-containing Species to the Composition and Optical Properties of DOM from Various Aquatic Locales 94
6.1 Introduction 94
6.2 Material and Methods 95
6.2.1 Reagents and Materials. 95
6.2.2 DOM Samples and Locations for Comparison 96
6.2.3 Sample Collection, Extraction, and Preparation 98
6.2.4 Mass Spectrometric Data Acquisition, Pre-processing, Molecular Formula Assignments, and Identification of Reducible Species in SRFA and DOM Extracts 99
6.2.5 Optical Measurements for SRFA. 99
6.2.6 Cluster Analysis for Similarity Comparison 100
6.3 Results and Discussion 100
6.3.1 Comparison of general mass spectrometric features. 100
6.3.2 Comparison of Molecular Formula Assignments 104
6.3.3 Comparison of Untreated and Reduced species 105
6.3.4 Similarities Between Samples from Different Aquatic Environments 120
6.4 Conclusions 123

7 Conclusions and Future Work 125
7.1 Conclusions 125
7.2 Future Work 127

A Internal calibrants and extraction blank 128

B Isobaric Species 130
B.1 Isobaric species: reduction with low and high amounts of $NaBD_4$ . . . 130

C Internal consistency 132
C.1 Calculating %I with three Methods A, B, and C 132

D Methods for Optical Measurements 135
D.1 Apparatus for Optical Measurements 135
D.2 Optical Measurements 135

E Supplemental Information for Chapter 4 138
E.0.1 Comparison of pre and post-extracted SRFA (untreated and reduced) 138

F Supplemental Information for Chapter 5 142
F.1 Comparison of pre and post-extracted SRFA (untreated, irradiated, and reduced) . . . 144

G Supplemental Information for Chapter 6 148
G.1 Sample collection for and preparation of C_{18} extracts . . . 148
G.2 North Pacific Ocean and Equatorial Atlantic Ocean Samples . . . 151

H MATLAB functions and descriptions 158
H.1 Preprocessing of raw mass spectral data (F1_list) . . . 158
H.2 Molecular formula assignment (F2_form) . . . 162
H.3 Search (F3_search) . . . 168
H.4 Find Reduced peaks (F4_Rint) . . . 172
H.5 Compare Script . . . 176

I MATLAB scripts: Fluoroscence 178
I.1 Fluorescence scripts for data collected on FluoroMax-4 luminescence spectrometer and converting intensity into QS units . . . 178

Bibliography 183

List of Tables

1.1 Optical Property Indices for Characterization of CDOM 12

2.1 Elemental Compositions of reference material SRFA[a] 28
2.2 DOM mass spacing patterns observed in the ESI mass spectra 30

4.1 Mass spectral peak distributions. 49

5.1 Mass spectral peak distributions (200-700 m/z) 74
5.2 Mass spectral peak distributions after reduction (200-700 m/z) 86
5.3 $\%I_{total}$ vs percent absorbance loss due reduction 92

6.1 Sample locations and descriptions for comparison 97
6.2 Mass spectral peak distributions of untreated samples from different aquatic environments (200-600 m/z) 101

A.1 m/z and corresponding molecular formulae (of [M-H]- ions) used for post-acquisition internal calibration of the mass spectra. 128
A.2 m/z and tentative assignments of peaks present in the extraction blank which were removed from mass spectra 129

C.1 %I calculated by three methods (from 200-600 m/z). 133

F.1 Percent absorbance loss due to irridiation and reduction. 146

G.1 North Pacific Ocean Water Masses from Hernes et al. 2002 [129] . . . 150

List of Figures

1.1 Subcategories of NOM . . . 2
1.2 Major sub-fractions of NOM and separation process . . . 4
1.3 DOM/CDOM interactions with the aquatic environment . . . 6
1.4 Absorption spectra of samples from three different locales. . . . 8
1.5 Spectral Slope from different locales . . . 9
1.6 Spectral Slope of Irradiated and Reduced SRFA . . . 10
1.7 Corrected emission spectra (EEMS) for reference material SRFA. . . 11
1.8 Wavelength dependence of fluorescence quantum yield for Suwannee River Fulivic acid . . . 13
1.9 Electron interaction model . . . 16
1.10 Changes in optical properties following reduction . . . 17

2.1 Schematic of a negative ion mode ESI . . . 21
2.2 Mass distribution of (untreated) Suwannee River fulvic acid (negative ion ESI FT-ICR MS) . . . 26
2.3 Possible number of formulae with different CHONS inclusions . . . 29
2.4 Presentation of mass spectral data . . . 31
2.5 Van Krevelen plots of assigned molecular formulae in SRFA depicting reaction pathways . . . 32
2.6 Van Krevelen plots of assigned CHONS molecular formulae (200-600 m/z) . . . 33

3.1 Reduction of ketone/aldehyde species . . . 35
3.2 Mass labeling with sodium borodeuteride. . . . 41
3.3 Illustration of MATLAB algorithm (Method A-Formulae) . . . 42

4.1 Sample preparation of SRFA . . . 46
4.2 Bar graph of %TPI of UNT and R-50x assigned equivocal/unequivocal and D-containing/non-D-containing molecular formulas . . . 51
4.3 Mass spectra (expanded region 467-473 m/z) of UNT and reduced (R3x and R50x) SRFA at 467 (left; M), 470 (middle; M+3), and 473 (right; M+6) m/z. . . . 53
4.4 Negative ion ESI FT-ICR mass spectra (200-600 m/z) of peaks assigned CHO molecular formulae. . . . 54

4.5 %I_X extent of reduction vs increasing mass excess $NaBD_4$ (2 to 157-fold) added to reduced samples. . . . 56
4.6 %N_X extent of reduction vs increasing mass excess $NaBD_4$ (2 to 157-fold) added to reduced samples. . . . 57
4.7 Van Krevelen plots of CHO formulae identified in low (200-400 m/z) and high (400-600 m/z) mass ranges. . . . 59
4.8 Double Bond Equivalents (DBE) vs m/z of CHO formulae identified. 61
4.9 Absorption (prior to solid phase extraction-PPL) of untreated and 24 hour-reduced samples and the percent loss of absorption following reduction . . . 62
4.10 Corrected emission spectra of untreated and reduced . . . 62
4.11 Logarithmic relationship between absorbance measurements as well as mass spectrometric measurements vs increasing amounts $NaBD_4$. . 63
4.12 Percent of integrated absorbance loss upon reduction $NaBD_4$. . . . 64

5.1 Absorption of untreated and irradiated samples . . . 73
5.2 Negative ion ESI FT-ICR mass spectra (200-700 m/z) for UNT and IRR of peaks assigned CHO molecular formulae. . . . 75
5.3 DBE vs m/z of untreated and irradiated SRFA samples . . . 76
5.4 Relative abundance of DBE . . . 77
5.5 Van Krevelen plots of CHO formulae identified in untreated and irradiated SRFA . . . 78
5.6 Effect of irradiation on SRFA at nominal mass 442 m/z . . . 79
5.7 Negative ion ESI FT-ICR mass spectra and Van Krevelen plots comparing IRR320 and two North Pacific Ocean samples . . . 82
5.8 Van Krevelen plots comparing what is lost upon irradiation of SRFA and from surface to deep ocean water . . . 83
5.9 Absorption and percent of absorption loss for irradiated and irradiated *and* reduced SRFA samples compared to untreated SRFA . . . 85
5.10 Negative ion ESI FT-ICR mass spectra (200-700 m/z) of peaks assigned CHO molecular formulae for reduced samples . . . 87
5.11 %I_X extent of reduction vs R50x and irradiated samples. . . . 88
5.12 Van Krevelen plots of CHO formulae identified in irradiated, reduced, and irradiated and reduced SRFA . . . 89
5.13 Double Bond Equivalents (DBE) vs m/z of CHO formulae identified in irradiated, reduced, and irradiated and reduced SRFA . . . 90
5.14 Percent of integrated absorbance loss upon reduction with $NaBD_4$ relative to untreated and respective irradiated samples (at pH 7) across three wavelength ranges correlated to %I_{total}. . . . 92

6.1 World map of sample locations . . . 97
6.2 Negative ion ESI FT-ICR mass spectra (200-600 m/z) for comparison of samples from different aquatic environments used for this study . . 103
6.3 Absorption spectra of five of the untreated comparison samples . . . 106

6.4 Spectral slope from 275-295 nm and $E2/E3$ ratio for comparison untreated samples) . . . 106
6.5 Van Krevelen plots of CHO formulae identified in five untreated comparison samples . . . 107
6.6 Absorption for Delaware samples normalized at 250 nm . . . 108
6.7 Van Krevelen plots comparing the loss of CHO formula from the Delaware upper bay to the Delaware shelf break in three mass ranges (200-400 and 400-600 m/z . . . 109
6.8 Absorption of untreated and 24 hour-reduced samples for SRFA and DEL select transect samples and percent loss of absorption following reduction . . . 110
6.9 Absorption of untreated and 24 hour-reduced samples for select NPO and EAO samples and percent loss of absorption following reduction . 111
6.10 Spectral slope from 275-295 nm and $E2/E3$ ratio for reduced comparison samples . . . 112
6.11 Negative ion ESI FT-ICR mass spectra (200-600 m/z) of peaks assigned CHO molecular formulae showing the reduction for eight samples from different aquatic environments . . . 114
6.12 $\%I_X$ extent of reduction vs different aquatic environments. . . . 115
6.13 Van Krevelen plots of CHO formulae identified in the low (200-400 m/z) mass range showing the reduction of eight samples from different aquatic environments . . . 117
6.14 Van Krevelen plots of CHO formulae identified in the high (400-600 m/z) mass range showing the reduction of eight samples from different aquatic environments . . . 118
6.15 Percent of integrated absorption loss upon reduction with $NaBD_4$ for the eight comparison samples (relative to the respective untreated sample) correlated to $\%I_{total}$ . . . 120
6.16 Heatmaps showing the similarities (based on Bray-Curtis analysis) of the CHO molecular formula in eight untreated samples from different aquatic locations (1,900 of the highest intensity peaks within 200-600 m/z) . . . 122
6.17 Heatmaps showing the similarities (based on Bray-Curtis analysis) of the reduced CHO molecular formula in eight samples from different aquatic locations (800 of the highest intensity reduced peaks from 200-600 m/z) . . . 123

A.1 Negative ion ESI FT-ICR mass spectrum of the extraction blank . . . 129

B.1 Potential reductions of possible isomers for a hypothetical DOM molecule 131

C.1 Illustration of MATLAB algorithms (Methods B and C) . . . 134

E.1 Absorption spectra at pH 7 before and after solid-phase (PPL) extraction. . . . 139

E.2 Negative ion ESI FT-ICR mass spectra of all samples with CHO-only assigned molecular formulae peaks 200-600m/z . . . 140
E.3 Optical properties of Pre and Post at pH7 correlated to $\%I_{total}$ (200-600 m/z) . . . 140
E.4 Percent of integrated absorbance loss upon reduction with NaBD4 relative to untreated (UNT) sample (pH7) across three wavelength ranges correlated to $\%I_{1O}$, $\%I_{1B}$, $\%I_{2B}$, and $\%I_{total}$. . . . 141

F.1 Block diagram of the sample preparation of untreated, irradiated and reduced SRFA. . . . 142
F.2 Absorbance of 418 nm and 320 nm long pass filters . . . 142
F.3 Absorbance over time for 418 nm and 320 nm long pass filters . . . 143
F.4 Irradiation setup . . . 143
F.5 Absorption spectra before and after solid-phase extraction for UNT and IRR samples . . . 144
F.6 Bar graph of %TPI of untreated, irradiated, and reduced samples assigned equivocal/unequivocal and D-containing/non-D-containing molecular formulas . . . 145
F.7 Increase in peak intensity after irradiation . . . 146
F.8 Realative abundance of DBE-O of untreated and irradiated SRFA samples . . . 146
F.9 Van Krevelen plots of 2B CHO formulae identified in irradiated, reduced, and irradiated and reduced SRFA . . . 147

G.1 Flow cart of the collection of natural water samples in the field and preparation . . . 148
G.2 Absorption of NPO and EAO samples . . . 151
G.3 $E2/E3$ ratio for untreated and $NaBD_4$ reduced samples for NPO depth profile and EAO samples . . . 152
G.4 Absorption loss after reduction by $NaBD_4$ for NPO and EAO samples 153
G.5 $\%I_X$ extent of reduction vs NPO depth profile and EAO . . . 154
G.6 Bar graph of %TPI for seven samples assigned equivocal/unequivocal and D-containing/non-D-containing molecular formulas . . . 155
G.7 $\%I_{total}$ vs CHO and $CHON_1S_1$ for different locals . . . 156
G.8 Van Krevelen plot of the untreated molecular formulae that all eight samples have in common . . . 156
G.9 Van Krevelen plot of the reduced molecular formulae all eight samples have in common . . . 157

List of Abbreviations

$A(\lambda)$	absorbance at wavelength (λ)
AI	aromaticity index
APCI	Atmospheric Pressure Chemical Ionization
APPI	Atomospheric Pressure Photoionization
CDOM	chromophoric dissolved organic matter
DBE	double bond equivalents
DEL U Bay	Delaware Upper Bay
DEL M Bay	Delaware Lower Bay
DOM	dissolved organic matter
EAO	Equatorial Atlantic Ocean
EI	electronic interaction (model)
EEMS	excitation/emission matrix spectra
ESI	electrospray ionization
FT-ICR	Fourier transform ion cyclotron resonance
FDOM	fluorescent dissolved organic matter
IHSS	International Humic Substances Society
IR	Infrared
IRR	Irradiated
MALDI	matrix assisted laser desorption ionization
MW_N	number- averaged molecular weight
MW_W	weight- averaged molecular weight
MS	mass spectrometry
m/z	mass to charge ratio
NOM	natural organic matter
NPO	North Pacific Ocean
NMR	nuclear magnetic resonance
POM	particulate organic matter
SPE	solid phase extraction
SRFA	Suwannee River Fulvic Acid
QSU	quinine sulfate units
RED	reduced
UV-vis	ultraviolet-visible
UNT	untreated sample

Chapter 1: Aquatic Natural Organic Matter (NOM): Background and Environmental Importance

1.1 Aquatic Natural Organic Matter (NOM) and Sub-Fractions

Complex heterogeneous ensembles of organic compounds are found virtually within all aquatic environments. These ensembles, as a whole, are referred to as natural organic matter (NOM); they consist of decaying material from plants and animals as well as their degradation products. NOM is an important component of natural waters, varying in concentration, complexity, and reactivity from one location to another as well as with the depth within the water column. The reactivity of NOM and its role in the global carbon cycle are closely related to its composition and structure (molecular weight and functional groups) [1, 2]. NOM is known to affect many biogeochemical processes such as the transport and bioavailability of nutrients, trace metals [2, 3], and pollutants [4–6]. However, even after a century of ongoing research, efforts to fully understand NOM composition (source(s)/structure) and reactivity within aquatic environments remains inadequate due to its inherent complexity.

Many studies have examined the variability of NOM across different locales.

Since there have been observed differences in NOM depending on location, conceptually it can be useful to divide NOM into two subcategories: riverine and marine NOM. (Figure 1.1).

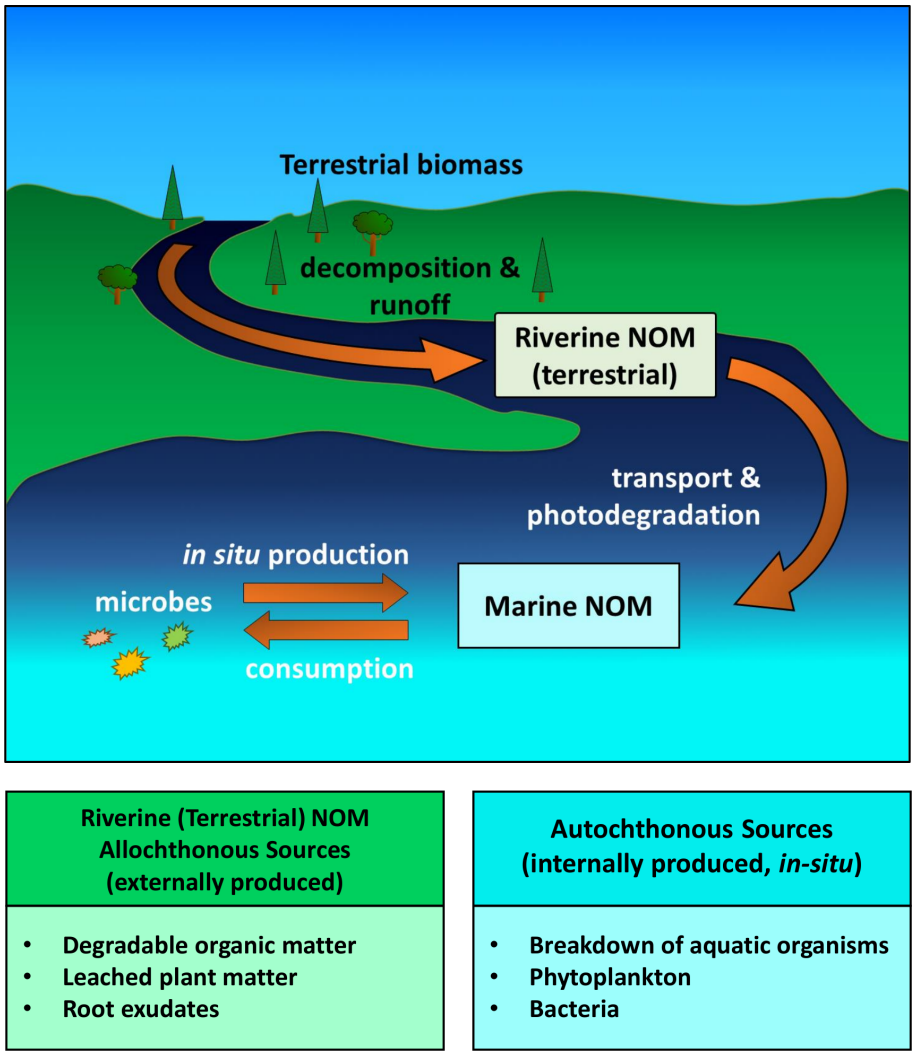


Figure 1.1: Top) Two location based subcategories of NOM (Riverine and Marine) and their main sources and sinks. Adpated from [7]. **Bottom)** Potential sources of marine NOM allochthonous and autochthonous. [8, 9]

Studies have provided evidence for riverine (terrestrially based) NOM to be primarily dominated by lignin degradation products [10–13] and possibly tannins

[10, 14, 15] (e.g. terrestrial plant and animal matter). These decomposition byproducts are then transported to larger bodies of water such as estuaries and lakes. However, in open oceans away from the influence of rivers, the source(s) of a subset of NOM (chromophoric dissoved organic matter (CDOM) which is discussed below) remains controversial; thus making the origin of marine NOM more ambiguous. Some researchers have provided evidence that marine CDOM is influenced by terrestrial sources [10, 13, 16], while others have suggested that marine CDOM arises *in-situ* from marine biomass [17] such as algae and phytoplankton [17, 18]; and others have proposed marine CDOM may act a food source for marine organisms [19]. Though in most cases, it is likely marine NOM is from a combination of both allochthonous and autochthonous factors: riverine NOM that has been degraded and/or photochemically altered and from the *in-situ* production and consumption from microbes (Figure 1.1) [8].

The factors/source(s) that contribute to marine NOM are important when trying to understand and predict its impact (reactivity and dynamics) within aquatic environments. Thus, knowing the molecular composition of NOM is needed. There were early attempts to categorize different compound types in NOM, by dividing NOM into sub-fractions using physical and chemical techniques. This doesn't necessarily separate NOM based on compound classes- but these separation techniques used are still used today to operationally define different sub-fractions of NOM (Figure 1.2). There are distinct differences in size between particulate organic matter (POM) and dissolved organic matter (DOM), as DOM is the material that can pass through a 0.2 or 0.45 μm filter, while the POM is the material that remains retained

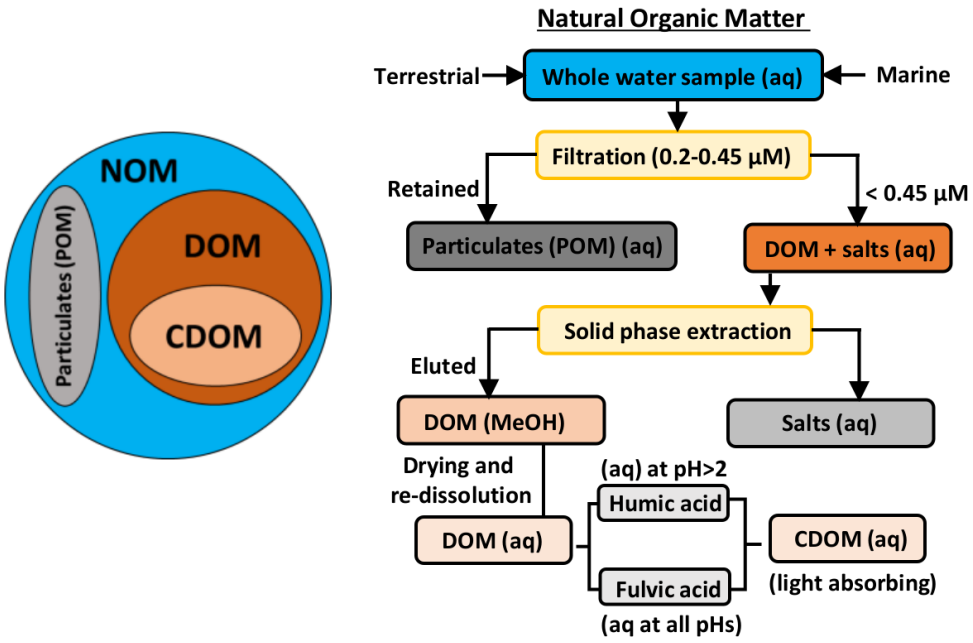


Figure 1.2: Left) Major sub-fractions of NOM and **Right)** General separation process for isolating these fractions from the aquatic environment and relationship between these fractions (NOM/POM/DOM/C-DOM) [20].

on the filter [20]. DOM can be further separated from inorganic compounds using different techniques, the most commonly used one is solid phase extraction (SPE). Furthermore, there is an optically active portion of DOM that can alter the aquatic light field due to its ability to absorb light in both the ultraviolet (UV) and visible wavelengths, called chromophoric dissolved organic matter (CDOM). This portion is not (currently) able to be separated from DOM but is the portion responsible for the photophysical and photochemical properties of DOM (Figure 1.2 and discussed in section 1.2 below).

In addition to these sub-fractions, there are sub-categories of NOM that make up humic substances (e.g. humin, humic acid, and fulvic acid), which are defined

based on their pH dependent solubility in water. Humin is not soluble at any pH and is therefore not considered a part of DOM [21]. Humic acid and fulvic acid are considered part of DOM as both are water soluble; but fulvic acid is soluble at all pHs, while humic acid is only soluble at pH>2 [21, 22].

Most research is conducted on sub-fractions, in particular DOM, of NOM using SPE techniques. These approaches allow for the consolidation of the organic material especially from areas where the amount of material may be very low, such as the open ocean. However, these separation techniques can lead to the preferential extraction of compounds (as well as the potential degradation of some compounds), potentially biasing the properties of the extracts relative to the bulk NOM/DOM [23]. These extracts are thus a fraction of the total NOM/DOM, representing only the material retained and eluted from the SPE column. Therefore, these approaches need to be used with discretion [23–25], though SPE columns (octadecyl carbon chain (C18) silica based column and functionalized styrene-divinylbenzene polymer (PPL) column) have been shown to have extraction efficiencies that provide largely representative extracts compared to their respective natural waters [26, 27].

1.2 DOM and CDOM Significance in the Aquatic Environment

As mentioned above, research on NOM has been ongoing for over a century. Very early research referred to DOM as yellow organic substances in riverine environments and gelbstoffe in seawater [20]. Over the decades, efforts to isolate and characterize DOM have improved by operationally defining DOM and increasing

analytical techniques and types of analytical instruments used to separate and to study DOM (analytical techniques used to study and characterize DOM discussed in Chapter 2).

DOM, especially marine DOM, is of ongoing interest because marine DOM is one of Earth's largest carbon reservoirs; the oceans contain almost as much dissolved carbon (∼700 Pg) as carbon dioxide (∼750 Pg) in the atmosphere [20,28,29]. Thus the potential impacts of marine DOM on the global carbon cycle are of importance [20]. Specifically, CDOM, the optically active portion of DOM, can have a considerable impact on important biogeochemical processes within the aquatic environment due to its ability to absorb light and undergo photochemistry [30–32] (Figure 1.3).

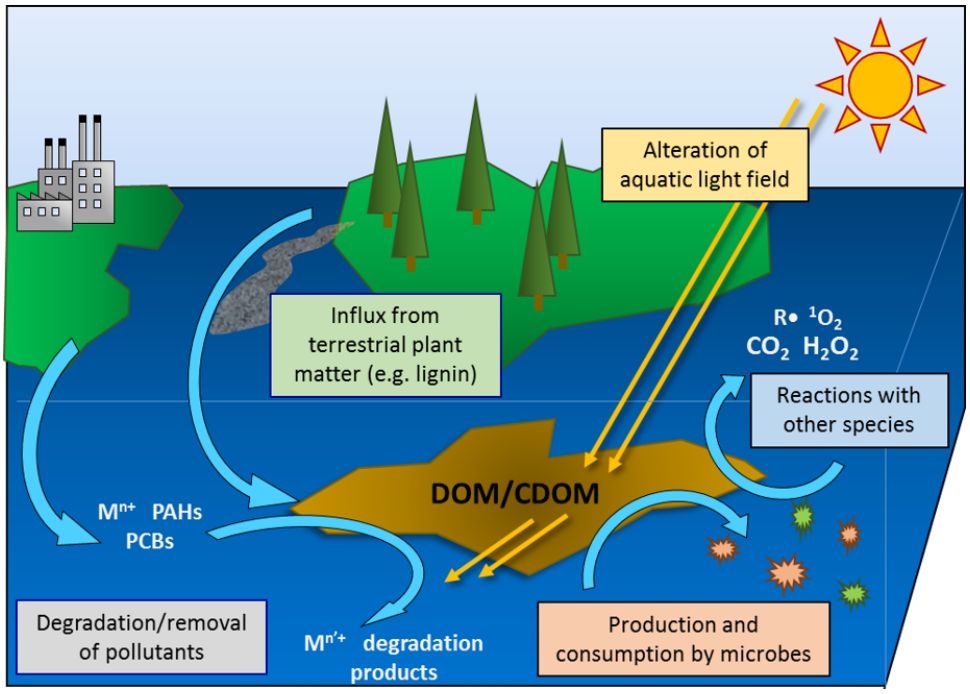


Figure 1.3: DOM/CDOM interactions with the aquatic environment [5,6,25,30,33]. Adapted from www.visionlearning.com.

For example, absorbing strongly in the UV, CDOM has the potential to significantly limit the penetration of harmful UV-B radiation, thus, protecting light-sensitive organisms in the aquatic environment [33], while absorbing in the visible wavelength region can decrease the amount of solar radiation accessible to phytoplankton, in turn potentially slowing photosynthesis [33]. Furthermore, light absorption by CDOM can initiate photochemical reactions that can generate inorganic species (e.g. carbon dioxide and carbonyl sulfide) [25], photosensitize reactions with other species in the aquatic environment (pollutants) [5], and can create reactive oxygen species (such as singlet dioxygen, hydroxyl radicals, superoxide radicals) [6]. The composition of DOM/CDOM determines its relative reactivity which in turn can control its residence time and fate in the environment [29], thus making it an influential part of the complex web that makes up the global carbon cycle. Determining the chemical structures that contribute to the optical properties of CDOM would help to understand and predict the role of CDOM in aquatic systems.

1.3 Optical Properties of CDOM

1.3.1 CDOM absorption and emission

To study the optical properties of CDOM, absorption and fluorescence spectroscopy are commonly employed. Both the absorption and emission of light by CDOM from aquatic environments follow very similar trends, even though there are many possibilities in the variability of composition due to the complexity of the ensembles. The CDOM absorption spectra are relatively featureless (rarely have any

discernible shoulders) and decrease exponentially with increasing wavelength from the ultraviolet (UV) to longer visible wavelengths [34, 35] (Figure 1.4).

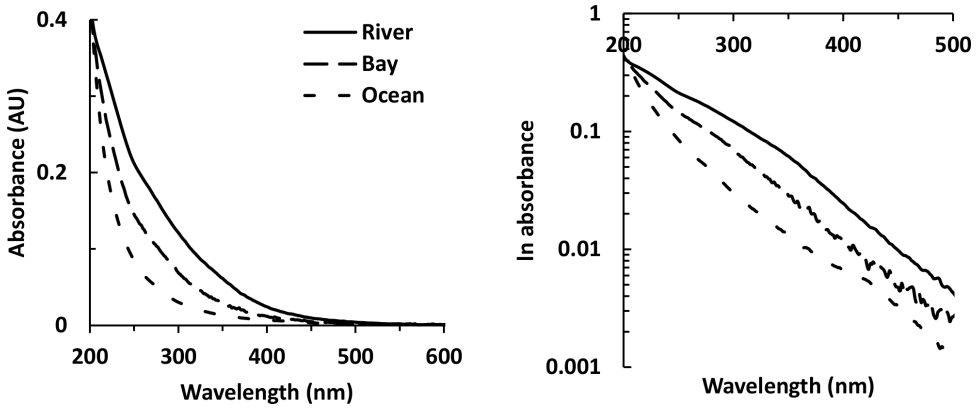


Figure 1.4: Absorption spectra of samples from different locales normalized at 200 nm. River- reference material Suwannee River Fulvic acid; Bay- C_{18} extract from the Delaware upper bay surface (2 m); and Ocean-C_{18} extract from the Equatorial Atlantic Ocean (1000 m). Absorption collected on a Shimadzu UVPC 2401 spectrophotometer using a 1 cm quartz cuvette at pH 7.

Due to the approximately exponential decrease in absorption with increasing wavelength, experimental absorption spectra are commonly fit to the equation

$$A(\lambda) = A(\lambda_{ref})e^{-S(\lambda-\lambda_{ref})} \tag{1.1a}$$

where $A(\lambda)$ is the absorbance at wavelength λ between a set range of wavelengths, λ_{ref} is a reference wavelength, and S is the absorption spectral slope, which is determined by how readily the absorption decays [35]. Due to changes in absorbance (and therefore spectral slope) of CDOM from different locales, the spectral slope can be used as a parameter for CDOM characterization and can potentially provide

useful information about the source of CDOM.

Studies have shown that fresh waters containing terrestrial CDOM and coastal waters that are influenced by large fresh water inputs have smaller spectral slopes relative to seawaters (marine CDOM) [36, 37] (Figure 1.5). This is partially due to terrestrial influenced CDOM having higher absorption at longer wavelengths, suggesting a higher degree of aromaticity and complexity than the marine CDOM [38].

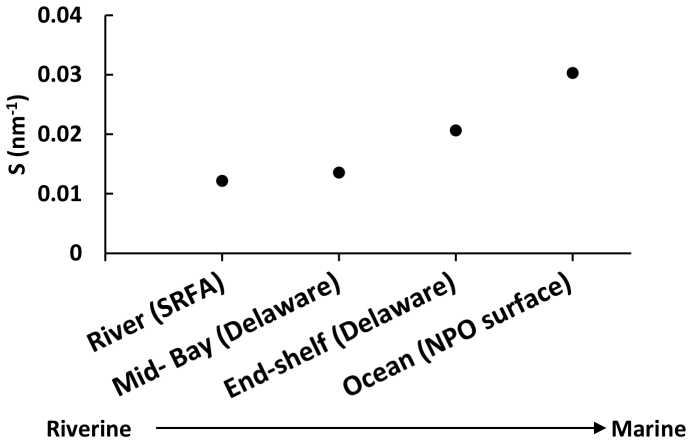


Figure 1.5: Spectral slope from 275-295 nm ($S_{275-295}$) for samples from different locales. River- reference material Suwannee River Fulvic acid; Mid-Bay- C_{18} extract from the Delaware bay surface (2 m); and Ocean- C_{18} extract from the North Pacific Ocean (surface-15 m). Absorption collected on a Shimadzu UVPC 2401 spectrophotometer using a 1 cm quartz cuvette at pH 7 used for spectral slope calculation.

While spectral slope does vary with source as shown above, it is important to note that it can also be altered through biological [16] and chemical processing (e.g. solar radiation/photobleaching and reduction, Figure 1.6) of the source material [30, 39, 40] or even mixing [38]. Other absorbance parameters, mainly specific wavelengths and ratios of wavelengths, have also been used as indexes for

characterizing CDOM (Table 1.1). However, it should be noted that even though there are correlations between these specific wavelengths (and ratios of wavelengths) to size/structure, the values can be affected by peaks/shoulders in the absorption spectra. Therefore, these indexes should be used with caution and a more complete analysis of the optical properties should be conducted.

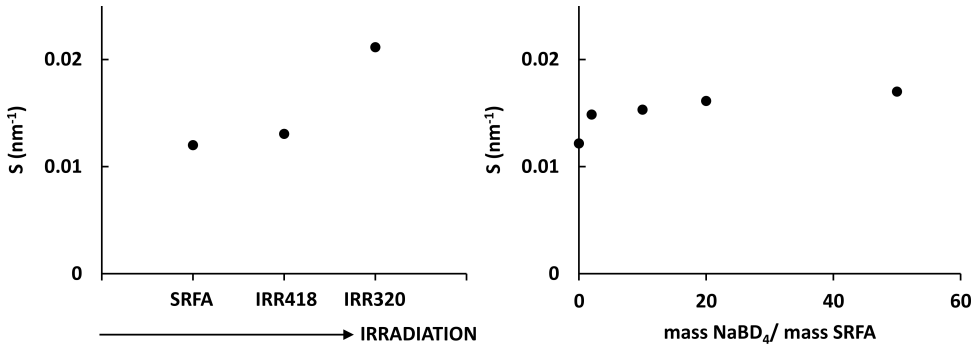


Figure 1.6: Spectral slope from 275-295 nm ($S_{275-295}$) **Left)** Untreated reference material SRFA and irradiated SRFA samples (IRR418- irradiated for 54 hours employing a 418 nm long pass filter; IRR320- irradiated for 33 hours employing a 320 nm long pass filter). Irradiated samples discussed in Chapter 5. **Right)** SRFA reduced with increasing amounts of sodium borodeutride (2 to 50 fold mass $NaBD_4$/mass SRFA). SRFA treated with increasing amounts of $NaBD_4$ discussed in Chapter 4.

CDOM also emits fluorescence over a broad range of excitation wavelengths, and is often denoted as fluorescent DOM (FDOM). Similar to the absorption spectra, the emission spectra of FDOM are relatively featureless; they are broad and unstructured [41] (Figure 1.7). Due to the complexity of DOM, excitation/emission matrix (EEMs) spectroscopy is usually employed to identify groups with fluorescent components within CDOM and thus provide more detailed information about the composition and structure [42, 43].

To collect EEMs, a sample is excited at different wavelengths and the emission spectrum is collected at each successively longer excitation wavelength. These emission spectra are then combined in one plot to visualize the fluorescence intensity over a range of excitations (Figure 1.7). The intensity of the emission exhibits a continuous red-shift in its maxima with increasing excitation wavelength, while overall decreasing in intensity (after excitation at 300 nm) (arrow and inset Figure 1.7) [10, 41, 44].

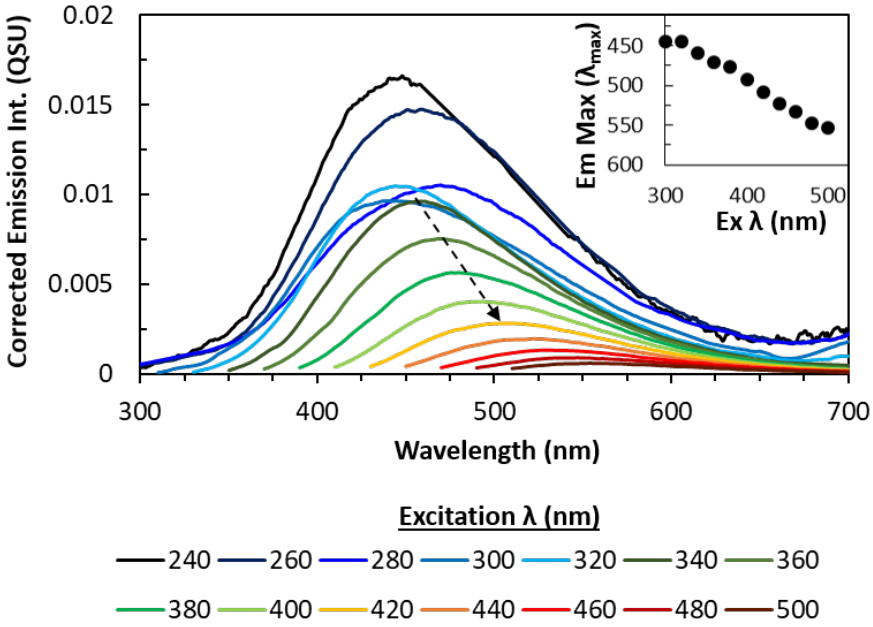


Figure 1.7: Corrected emission spectra (EEMS) for reference material Suwannee River Fulivic acid (excitation 240-500 every 10 nm over 300-700 nm). Emission intensities in quinine sulfate units (QSU). Inset: Shifting wavelength emission maximum at excitation wavelengths 300-500 nm (every 10nm). Collected on a Fluoro-Max 4 luminescence spectrometer using a 1 cm quartz cuvette at pH 7.

A few specific, excitation emission wavelengths have been used as indexes for characterizing FDOM (Table 1.1). Once again, as with the absorbance, these correlations (this time to the "type" of DOM), should be used with caution and in

conjunction with a more complete analysis of the full EEMs to gather information on fluorescent components as well as structure and source.

Table 1.1: Optical Property Indexes for Characterization of CDOM (adapted from Hautala *et al.* 2000 [45] and Coble *et al.* 1996 [42])

	Wavelength (nm)	**Correlation/Characteristic** [42, 45–49]
Abs.	272, 280	Molecular weight, Aromaticity
	250/365 (E_2/E_3)	Molecular weight, Aromaticity
	465/665 (E_4/E_6)	Molecular weight, Humification, condensation of aromatic carbon
Fluor.	EX 275; EM 310	protein-like (tyrosine)
	EX 275; EM 340	protein-like (tryptophan)
	EX 400; EM 660	pigment-like (chlorophyll)
	EX 260; EM 380-460	UVC humic-like
	EX 290-310; EM 370-410	UVA marine humic-like
	EX 320-360; EM 420-460	UVA humic-like

1.3.2 Quantum Yields

Fluorescence quantum yields (QY or ϕ), the ratio of photons emitted to those absorbed, can used for analyzing CDOM from different locales. This ratio is compared to a known standard, most commonly quinine sulfate, as seen in equation

$$\phi(\lambda) = [F'(\lambda) * a_{QS}\phi_{QS}]/[a(\lambda) * F'_{QS})] \tag{1.2a}$$

where F' is the integrated fluorescence emission of the sample and F'$_{QS}$ is the integrated emission of quinine sulfate at excitation wavelength (λ); $a(\lambda)$ and a_{QS} is the absorbance of the sample and quinine sulfate, respectively, at the excitation wave-

length λ (the published value for the reference quinine sulfate (ϕ_{QS}) is 0.51). [34]

Quantum yield values are usually relatively low (0.001-0.03), have maximum values which commonly arise from excitation wavelengths between 350 and 450 nm, and values that decrease at both shorter and longer excitation wavelengths (Figure 1.8). The peaks in the QY that arise with excitation wavelength can also provide information about the source and structure of CDOM. [10, 34, 40, 44, 50]

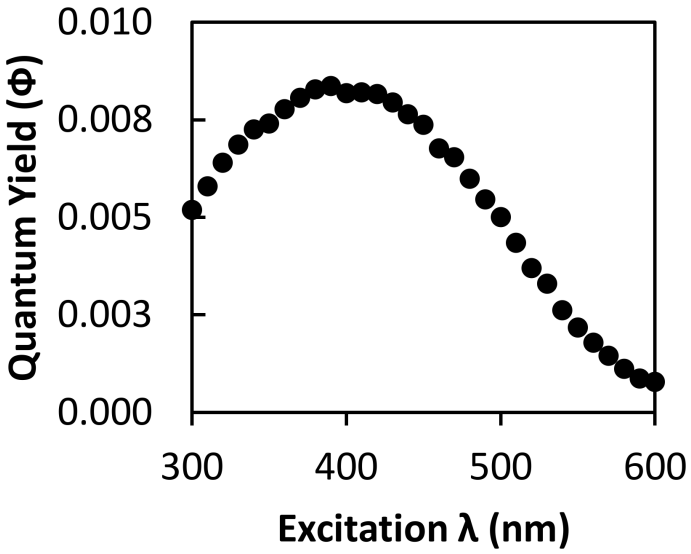


Figure 1.8: Wavelength dependence of fluorescent quantum yield for reference material Suwannee River Fulivic acid. Absorption collected on a Shimadzu UVPC 2401 spectrophotometer using a 1 cm quartz cuvette and fluorescence collected on Fluoro-Max 4 luminescence spectrometer using a 1 cm quartz cuvette (pH 7).

1.3.3 Structural Basis of the Optical Properties of CDOM

Several analytical techniques, such as infrared (IR), nuclear magnetic resonance (NMR), mass spectrometry (MS), and electrochemical techniques have been employed to gain insight into the structural basis of DOM/CDOM (which will be described in Chapter 2). Through the use of these techniques, carboxyl and pheno-

lic functional groups have been shown to be the dominant functional groups within DOM samples [51–54]. Carbonyl-containing functional groups such as aromatic ketones, aldehydes, and quinones have also been identified [55, 56]. Though, due to the complexity and variability of CDOM in aquatic environments, relating the bulk observable optical properties (e.g. light absorption, fluorescence emission) to its structure and composition has proven to be difficult.

To explain the optical properties of CDOM, two models have been proposed: 1) superposition model and 2) electronic interaction (EI) model. In the simple superposition model, absorption and emission are said to arise from independently (non-interacting) absorbing and emitting chromophores within the DOM sample [40]. Most organic compounds absorb at wavelengths shorter than the visible spectrum (>400nm), with the exception of highly conjugated aromatic species and/or quinones (which can have have n $\rightarrow \pi^*$ transitions that occur out into the visible wavelength region). Though, this longer wavelength absorption in the visible is usually weak (example model compound Vanillin Figure 1.9-Bottom).

This leads to, as many studies have demonstrated, that this superposition model cannot readily explain the long wavelength absorption or the complex spectral dependence of the steady-state and time resolved emission of the photochemical behavior of CDOM [41, 44, 57].

In contrast, the EI model does account for CDOM optical properties; thus, it is likely that many of the unique optical properties of these complex ensembles, particularly those of long wavelength absorption in the visible, are influenced by chromophore-chromophore interactions between polyhydroxy/methoxy aromatic

electron donors (e.g. phenols) and carbonyl-containing electron acceptors (e.g. aromatic ketones/aldehydes) within CDOM [40, 41, 58, 59].

Though the EI model can account for many of the optical and photochemical properties of CDOM, additional tests are ongoing to better understand the structures responsible for the optical properties. Such tests include the selective chemical reduction of specific functional groups with sodium borohydride and sodium borodeuteride ($NaBD_4$), the pH titration, irradiation experiments, and mass spectrometry experiments.

Specifically, sodium borohydride/deuteride is known to selectively and irreversibly reduce ketone and aldehyde containing species [10,57,60–65]. The reduction of these ketone/aldehyde species (acting as acceptor moieties in DOM) in turn should result in the interruption of electronic interactions/disruptions of the donor/acceptor complexes and thus the loss of long wavelength absorption and emission (Figure 1.9). Experiments using $NaBH_4$ and $NaBD_4$ to reduce CDOM, have shown the loss of absorbance [10, 57, 60, 64, 65] (with greater fractional loss at successively longer wavelengths, also shown in Figure 1.10).

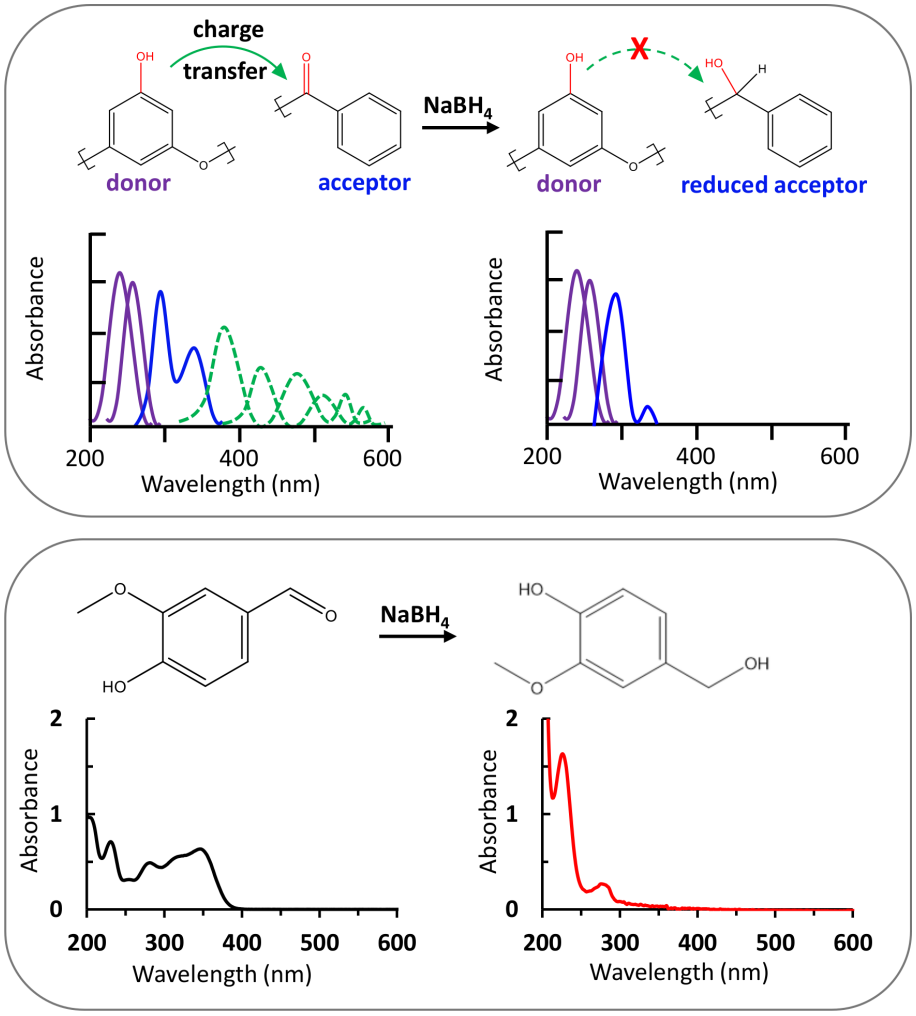


Figure 1.9: Electron interaction model: **Top)** Schematic of donor/acceptor moieties and reduction with $NaBH_4$. Depiction of a potential absorption spectrum of CDOM before (left) and after (right) treatment by $NaBH_4$ showing the loss of longer wavelength absorption when EI are disrupted by the reduction of the acceptor moieties. **Bottom)** Model compound (Vanillin) absorbance from before (black line) and after (red line) treatment with $NaBH_4$ showing an example of the absorbance loss in the longer wavelength region as described in the schematic above. Absorption collected on a Shimadzu UVPC 2401 spectrophotometer using a 1 cm quartz cuvette.

Furthermore, local donor moieties are no longer quenched due to the reduction of these carbonyl acceptor moieties; this should result in an enhanced blue-shifted

emission, which is observed in the fluorescence experiments [10, 60] (Figure 1.10).

As stated above the loss of this long wavelength absorption would be difficult to explain by the superposition model only (due to isolated ketones and aldehydes usually not absorbing at wavelengths greater than 400 nm). Therefore, these interactions suggest the long wavelength absorption is due to the electronic interactions between donors/acceptors.

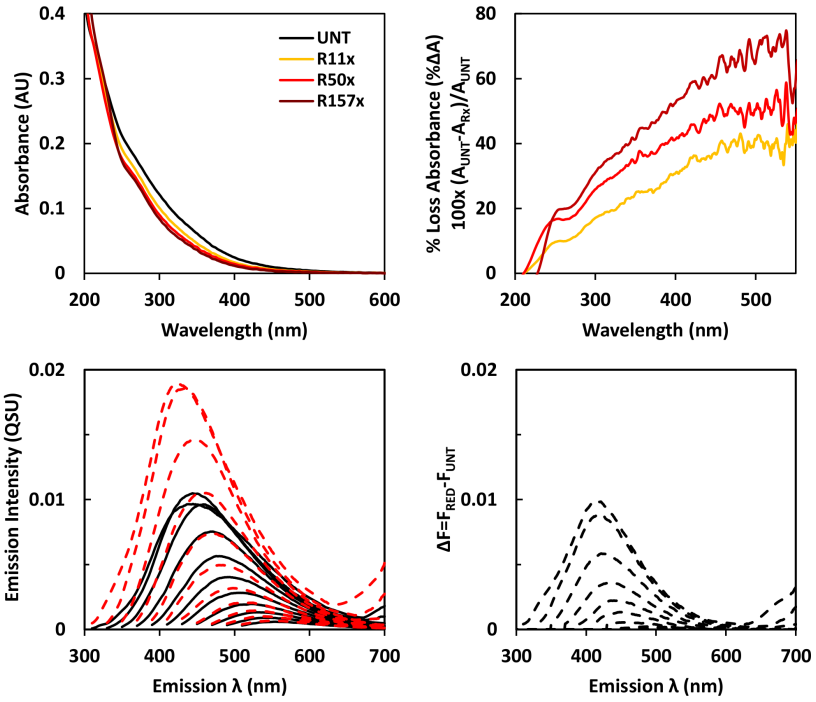


Figure 1.10: Changes in absorption and emission of SRFA following reduction with increasing amounts of borodeuteride (11, 50, and 157-fold $NaBD_4$ mass excess/mass SRFA respectively) after 24 hours. **Top)**Left- Absorption of untreated (UNT) and reduced (Rx) samples. Right- Percent loss of absorption (relative to UNT) following reduction with increasing amounts of borodeuteride. **Bottom)** Corrected Fluorescence emission spectra of untreated (black) and $NaBD_4$ reduced, R50x, (red) SRFA in quinine sulfate units (QSU), ΔF (dashed black) shows the gain in fluorescence intensity resulting from reduction. Fluorescence collected on Fluoro-Max 4 luminescence spectrometer using a 1 cm quartz cuvette at pH 7.

Chapter 2: Analytical Techniques Used to Study and Characterize DOM: Focus on Ultrahigh Mass Spectrometry

2.1 Early studies: Structure and Composition of DOM

The early research on DOM (humic substances) tried to describe the structure of an "average" DOM molecule [66, 67]. However, this description of an "average" DOM molecule and its hypothetical structure can be misleading because DOM is known to consist of hundreds to thousands of distinct chemical species (that can vary from source to source). Most likely there is not a typical or average DOM molecule that exists in any given DOM sample. Furthermore, some early proposed models did not include aromatic moieties that are capable of absorbing in the UVA and visible light; and thus are not able to explain the long wavelength CDOM optical properties. Therefore, rather than thinking of an "average DOM molecule," former colleague Daniel Baluha suggested it could be better to think and describe a complex assortment of classes of compounds that could exist (within a DOM ensemble- e.g. lipids, peptides, cellulose, condensed hydrocarbons, lignin, condensed tannins) [7].

2.2 Analytical Techniques for Analyzing DOM

Most analytical techniques, though they contribute valuable information to the field, are only able to measure bulk properties due to the inherent complexity of DOM. As discussed in Chapter 1, UV-Vis and fluorescence spectroscopy allow for the characterization of bulk optical properties of the portion of DOM that absorbs light, CDOM. Other methods such as nuclear magnetic resonance (NMR) and Fourier transform infrared (FTIR) spectroscopy have provided some characterization of bulk DOM; for example, they are able to measure relative abundance of structural features-functional groups/chemical classes. In most cases these techniques can give an idea of molecules that may be present but can not prove them, unless there is additional data or other analytical data to confirm. Though, there is also information to gather from what is *not* there (from the NMR chemical shifts or the IR chemical space); thus, blank areas can be used to conclude the absence of a type of molecule/functional group. These techniques are not able to provide much data related to the structure of specific compounds in DOM due to the vast number of chemical species that comprise DOM. [68–70]

However, mass spectrometry allows for more than just bulk properties of DOM to be studied as it can detect discrete species/fragments of species by their mass to charge ratio (m/z) due to the high resolution able to be achieved with the instruments [69]. Soft ionization, such as electrospray ionization (ESI), coupled with ultrahigh resolution Fourier transform ion cyclotron resonance (FT-ICR) has the ability ionize compounds without fragmentation allowing for the detection of dis-

crete chemical species within DOM [52].

2.3 Instrumentation for Mass Spectrometric Analysis of DOM

2.3.1 Electrospray Ionization of DOM

Over the past few decades, electrospray ionization in conjunction with mass spectrometry has allowed for great advances to be made using mass spectrometry to study DOM and has emerged as the ionization method of choice for DOM analysis. Its wide scale use is due to ESI being able to ionize a wide range of (nonvolatile) organic molecules with polar functional groups in water-based and methanolic solvents and well as ESI being a "soft" ionization technique which does not produce fragments of compounds. ESI can produce either positively or negatively charged ions and although, positive ion mode usually is able to achieve greater sensitivity, it can also lead to the formation of sodium adducts ($M+Na^+$), and therefore complicate the mass spectrum [71–74]; therefore, for the analysis of DOM negative ion mode is predominately used. The general scheme of negative ion mode ESI is shown in Figure 2.1.

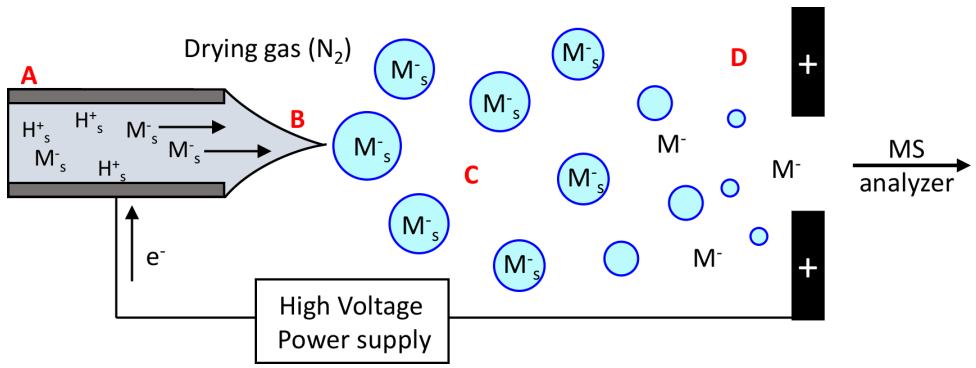


Figure 2.1: Schematic of a negative ion mode ESI. **A)** Electrospray capillary, **B)** Taylor cone, liquid filament forms, breaks up into droplets, **C)** anions migrate, **D)** solvent evaporates. [75–78]. Adapted from [7].

In ESI, a strong electric field is applied under atmospheric pressure and the sample is sprayed from a thin metallic capillary that has an onset voltage depending on the surface tension of the solvent (between 3-6 kV). Briefly, with increasing voltage, which in negative ion mode will be held at a negative potential, a charged droplet will form at the tip of the capillary from an extended liquid filament from what is called the Taylor cone [76]. Once the surface tension is broken, through increasing electrostatic static repulsion of the like charges in the droplet, the droplet will break into small charged droplets. This will cause these small charged droplets to drift away from each other and the solvent contained in them to evaporate, which in turn will cause the droplets to shrink and their charge (per unit volume) to increase. The droplets will also be drifting towards the (positively charged) counter electrode facilitated by the flow of nitrogen. This process will continue, small droplets breaking apart into smaller droplets, until gas phase ions are formed and drift all the way to the counter electrode located at the instrument inlet. [77,78]

The type of solvent used for ESI can significantly influence the efficiency of ionization. For example, water by itself is a poor solvent compared to methanol, acetonitrile, and dichloromethane [78]. And as mentioned above, one of the reasons ESI is used for DOM analysis is due to the compatibility of the solvent system. Since SPE procedures for DOM use methanol to elute the absorbed material, the eluate can be easily analyzed by ESI.

Despite the widespread use of ESI FT-ICR MS as a powerful tool for the characterization of DOM samples, ESI can only ionize a fraction of all species in a DOM sample. There are other ionization techniques that can be used such as atmospheric pressure chemical- or photoionization (APCI, and APPI, respectively) [78, 79] and matrix assisted laser desorption ionization (MALDI) [80, 81] that can ionize slightly different subsets of DOM, thus offering complementary compositional information.

2.3.2 Mass analyzers and ultrahigh resolution

Combining ESI with ultrahigh resolution FT-ICR MS has been used extensively to study the composition of DOM by resolving and assigning molecular formulae to thousands of species within a single DOM sample [64, 82, 83]. To be able to differentiate between compounds and thus be able to assign molecular formulae, the mass spectrometer must be able to resolve species with similar m/z values to be able differentiate between compounds. FT-ICR MS instruments have the highest resolving power (mass accuracy) among all mass spectrometers [84]. The resolving

power (RP) is the ability of the instrument to separate two distinct signals which have a small m/z difference and is usually defined by the equation

$$RP = \frac{M}{\Delta M} \tag{2.1a}$$

where M is the measured mass and ΔM is the width of the peak at half of the maximum intensity [77, 85]. Full width at half the height maximum (FWHM) of a specific ion (m/z value) has been accepted as the way to report RP. In DOM samples, m/z 400 is usually used for the FWHM. Furthermore, FT-ICR MS instruments can achieve RP >500,000 (FWHM), which corresponds to peak widths of 0.001 m/z resolvable at 500 m/z [77, 85]. Other mass spectrometers, such as time-of-flight and quadrupole, have lower resolving power and thus can only be used to gain insights into the general composition of DOM [85]. The high resolving power of the FT-ICR mass analyzer is what allows for the detection and formula assignments of thousands of species within DOM.

All FT-ICR MS instruments have four main components in common- strong magnet, analyzer cell (ICR), ultrahigh vacuum, and data processing system. The field strength ranges from 3 to 9.4 Telsa (T) for superconducting magnets and 20 T for resistive magnets; as the magnetic field increases so does the performance of the instrument. The analyzer cell, called a Penning trap in the case of FT-ICR MS, is where the ions are trapped and masses are analyzed and detected. Since FTMS is sensitive to pressure, an ultrahigh vacuum (10^{-9} - 10^{-10} Torr) is necessary to detect the ions (at a high resolution).

There are a few fundamental aspects of FT-ICR MS. First, the path of ions are curved in a magnetic field and if the field is strong enough, the ions will move in a circular motion perpendicular to that of the applied magnetic field and get trapped, hence the name Penning trap. The ions will rotate at frequency defined by this equation

$$\omega_c = \frac{q}{m}\beta \tag{2.2a}$$

where ω_c is the cyclotron frequency (in radians per second), β is the strength of the magnetic field, q is the charge of an ion and m is the mass of the ion ($m/q = m/z$) [77,84]. Since the magnetic field is uniform, β remains constant and therefore the m/z is determined by measuring its respective cyclotron frequency. Ion motion can be observable when an electric field (externally applied) has the same frequency as the as the ion cyclotron frequency applied. The electric field at the same frequency allows for resonance absorption of the ion and the energy transfered to the ion will increase its kinetic energy; which in turn will increase the ions' oscillating path. The signal from the cyclotron resonance results from the ion's motion (and is perpendicular to the trajectory of the ions). As the ion gets close to the two detection plates (cell wall) an alternating current is produced. To be able to detect the signal, ions of the same mass must be excited/accelerated to the same energy so that they are in the same circulating path with the same frequency. However, the ions must be be clustered together, which is done by only exciting the ions for a very short amount of time (to avoid having signals cancel out). After the ions

have been excited and the current created, the time-domain signal is converted to the frequency domain via Fourier transform. The alternating current produces a signal that varies sinusoidally with time which can be amplified and calibrated and converted into a mass spectrum. [77, 86, 87]

2.4 Mass spectral Analysis, Interpretation, and Representation of DOM

2.4.1 Mass Distribution

The ultrahigh resolution mass spectra of DOM samples contain several thousand peaks and have similar mass distributions. The peak distribution of DOM is centered around 300-500 m/z and each nominal mass consists of several peaks (Figure 2.2). The average molecular weights observed by ESI FT-ICR MS are lower than those observed by other methods [80], such as size exclusion chromatography or small angle X-ray scattering (500-4000 Da [46, 47] or 500-10,000 Da [88], respectively).

The lower molecular weights observed with ESI as compared to other analysis techniques has been attributed to two possible factors. First, ESI may not be able to efficiently ionize high molecular weight species and second, humic substances, a component of the DOM, are composed of supramolecular aggregates formed from lower molecular weight species. The aggregates would remain aggregated when analyzed by bulk methods but would be disrupted by the ESI process. [80, 89]

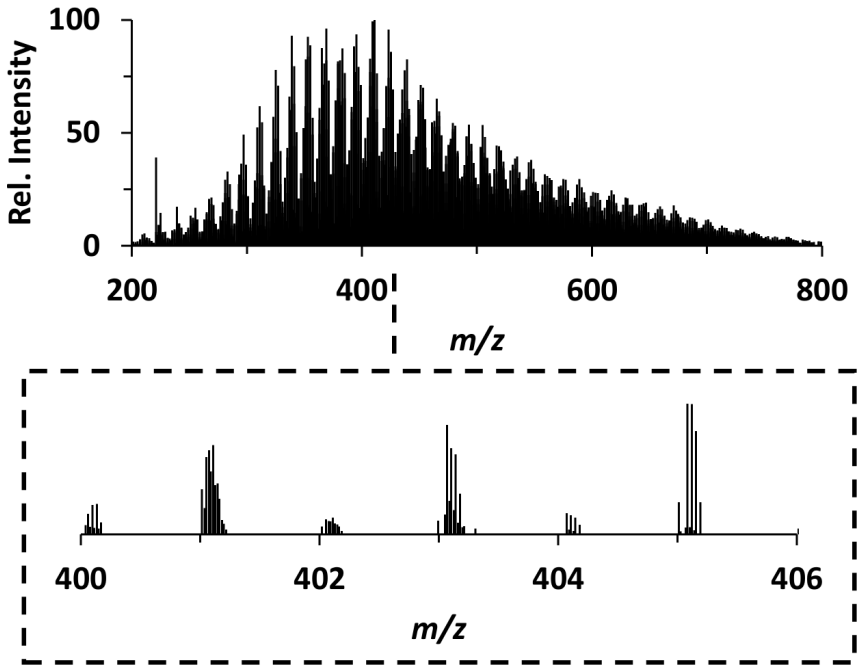


Figure 2.2: Mass distribution of (untreated) Suwannee River fulvic acid (negative ion ESI FT-ICR MS) of **Top)** full spectrum 200-800 m/z **Bottom)** Zoom from 400-406 m/z depicting multiple peaks at each nominal mass.

In addition, there are sample matrix effects that can contribute to differences in ionization efficiency which can lead to not all the components of the DOM sample being ionized and therefore not represented in the mass spectrum. Methods and parameters are thus currently optimized to give the highest sensitivity, resolution, and reproducibility at lower m/z ranges (>1000 m/z). [29]

2.4.2 Molecular Formula Assignment

As mentioned above, a typical DOM peak list from ESI FT-ICR MS contains thousands of peaks; to assign molecular formula to these peaks, the masses from the

peak list are first compared and converted to a list of measured neutral masses (the measured m/z of a singly-charged, deprotonated species with the addition of the exact mass of a proton). These converted masses are then assigned a formula through the use of software programs, such as MATLAB algorithms discussed in Chapter 3) and if the resulting formula error (FE) is less than the defined error cutoff, which is usually set at 1 ppm or less, reliable assignments are able to be obtained (the lower the FE is set, the more reliable the assignment is). The equation used to calculate the formula error is

$$FE = 1,000,000 \times (M_{theor} - M_{meas})/M_{theor} \tag{2.3a}$$

$M_{thoretical}$ is the calculated mass of the assigned formula and $M_{measured}$ is the measured converted neutral mass of the peak. For analysis, the mass accuracy of the instrument will define how low the FE limit can be set. FT-ICR MS (and some orbitraps) can achieve mass accuracies of less than 1 ppm within the mass ranges typically used for DOM analysis. [15, 90]

In addition to FE, the number of elements and atoms of each element that are used to calculate the assignments can greatly affect the molecular formula that are able to be assigned. Allowing more heteroatoms in the assignments allows more peaks to be assigned; however, allowing more heteroatoms also increases the number of formulae that fall within a given FE window of a specific single measured mass. Since, it is known that DOM consists mostly of C, H, O, and N with minor contributions from S and P [91] (Table 2.1) these are the elements included in the

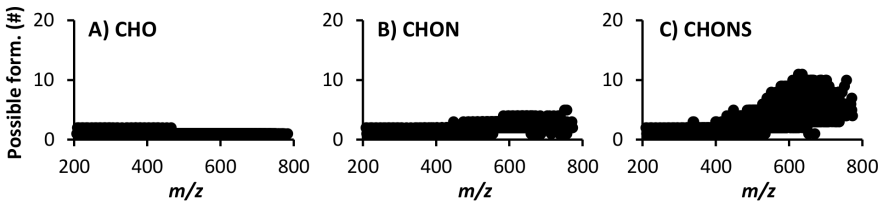
molecular formulae searches.

Table 2.1: Elemental Compositions of reference material SRFA[a]

	C	H	O	N	S	P
SRFA	52.34	4.36	42.98	0.67	0.46	0.004

[a]elemental composition in %(w/w) from IHSS (http://humic-substances.org/)

When more heteroatoms are included there could be 10 or more possible formulae that fall within a measured mass with an FE window of even <1 ppm, especially at higher masses [91] (Figure 2.3). In addition, when only $C_{0-\infty}H_{0-\infty}O_{0-\infty}$ formulae are considered usually >∼50% of the total peaks can be assigned formulae unequivocally. Although this extent of peak assignment might not sound significant, if one considers the total peak intensity able to be assigned, which is typically >80%, it can be seen that the majority of the sample can be assigned formulae [91] (Figure 2.3). Therefore, most studies only allow for small numbers of non-oxygenated heteroatoms, such as $N \leq 2$ or $S \leq 2$, to reduce the amount of possible molecular formulae for one given mass [82, 83].



	CHO	CHON	CHOS
Peaks assigned:	4856 (62%)	6268 (80%)	6953 (87%)
Possible formulae:	4951	11071	20963
Avg. form. per peak:	1	2	3
TPI assigned:	1.26E+11 (85%)	1.36E+11 (92%)	141.E+11 (96%)
Unequiv. assignments:	84%	44%	42%
Equiv. assignments:	1%	48%	54%

Figure 2.3: Top) Possible number of formulae with different CHONS inclusions for a negative ion ESI FT-ICR mass spectrum of (untreated) Suwannee River fulvic acid (total of 7841 peaks from 200-600 m/z). **Bottom)** Peaks assigned and (percentage of the total assigned), total number of possible formulae, average number of formulae per peak. **A)** $C_{0-\infty}H_{0-\infty}O_{0-\infty}$, **B)** $C_{0-\infty}H_{0-\infty}O_{0-\infty}N_{0-30}$, **B)** $C_{0-\infty}H_{0-\infty}O_{0-\infty}N_{0-30}S_{0-2}$. (Similar trends were seen in Koch et al., 2007 [91]).

When heteroatoms are included, steps need to be taken to assign the most reliable formula as well remove false positives or unreliable peaks that do get assigned. One of the most accepted methods chooses the formulae with the least amount of heteroatoms [92]. Also, the common mass spacing patterns found in ultrahigh resolution MS of DOM (Table 2.2) can also be used for assigning "unassigned" molecular formula. For example, formulae are first assigned based on the least amount of heteroatoms, then peaks that have not been assigned but differ from an assigned peak by one of the spacing patterns can be used to derive a molecular formula. This approach can be applied to identify isotopologues (e.g. ^{13}C and ^{34}S). Furthermore, a way to identify and then remove unreliable assignments, is by considering the

range DBE-O (double bond equivalents calculated minus the number of oxygen in the formula) of the assignment [93]. If the DBE-O of a formulae falls outside of the range of being reliable, the peak can be removed. [82, 94]

Table 2.2: DOM mass spacing patterns observed in the ESI mass spectra

***m/z* difference**	**Molecular formula difference**
+14.015 650	+ CH_2
+0.036 385	+ CH_4–O
+2.015 650	+ H_2
+0.995 249	+ N–CH
+1.003 355	+ ^{13}C–^{12}C
+1.995 769	+ ^{34}S–^{32}S
+0.003 371	+ $^{32}SH_4$ –$^{12}C_3$

2.4.3 Presentation of Mass Spectrometric Data

Due to the complexity of DOM, being composed of thousands of species and FT-ICR MS spectra containing thousands of peaks to which molecular formulae can be assigned, an extensive amount of data is acquired from each spectrum. To interpret the large amount of data collected, visual schemes are commonly used because the depiction of m/z against intensity alone does not allow for evaluation of the molecular complexity of DOM [52] (Figure 2.4-Top). Van Krevelen diagrams are used to depict H/C against O/C molar ratios of the calculated molecular formulae where each point represents the formula of a single peak (Figure 2.4-Middle). Other diagrams used to depict the data are carbon vs m/z, and double bond equivalents

(DBE) vs m/z (Figure 2.4-Bottom).

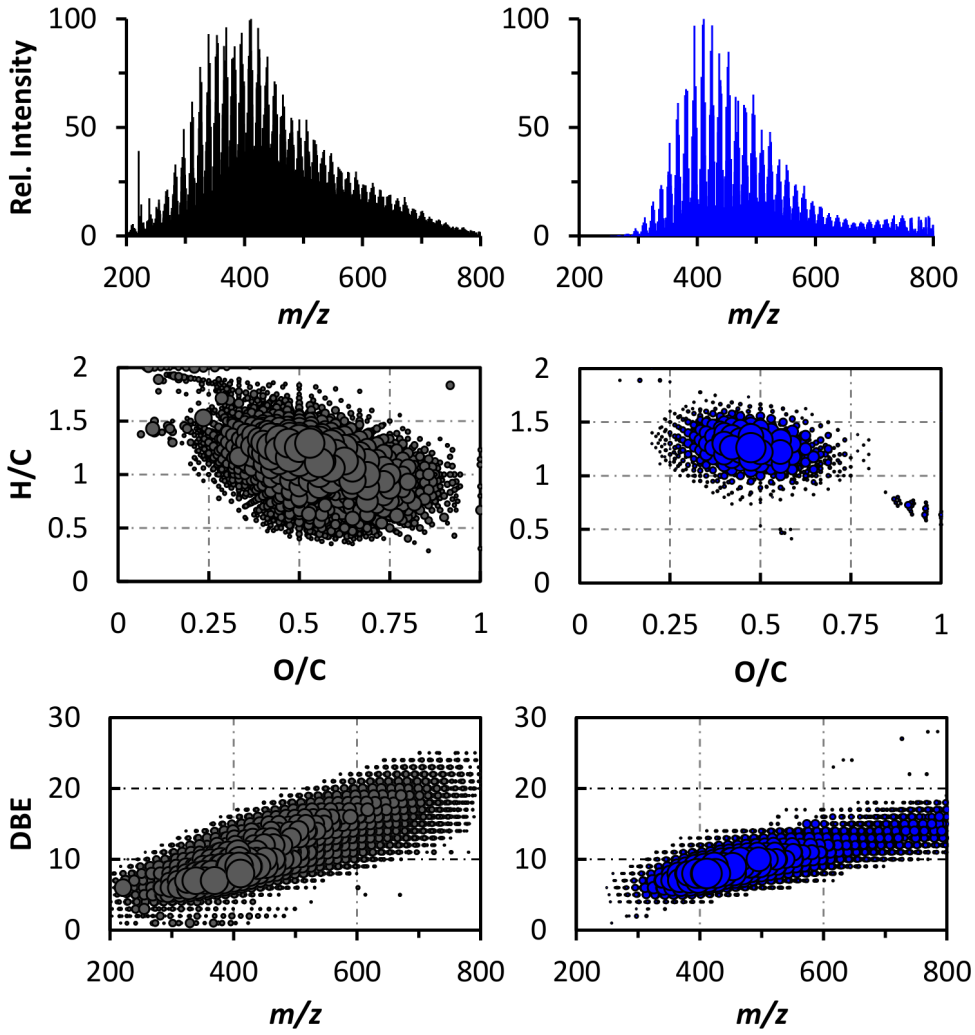


Figure 2.4: Presentation of mass spectral data from negative ion ESI FT-ICR mass spectra (200-800 m/z). **Top)** Relative intensity vs m/z. **Middle)** Van Krevelen plots of CHO formulae identified (200-800 m/z). **Bottom)** Double Bond Equivalents (DBE) vs m/z of CHO formulae identified (200-800 m/z). Untreated samples: River (SRFA) shown in black; Ocean (North Pacific Ocean depth of 1500 m) shown in blue.

Van Krevelen diagrams allow for observing possible compositional differences and/or similarities between samples. For example, in Figure 2.4 both SRFA and ocean samples have O/C ratios centered around 1; however, the SRFA sample has

H/C ratios extending below ∼1 and O/C ratios above ∼0.75, whereas the ocean samples do not. In addition, the chemical changes (reaction pathways) mentioned in Table 2.2, methylation/demethylation, hydrogenation/dehydrogenation, condensation/hydrolysis, oxidation/reduction, can be visually seen on the Van Krevlen diagram (Figure 2.5).

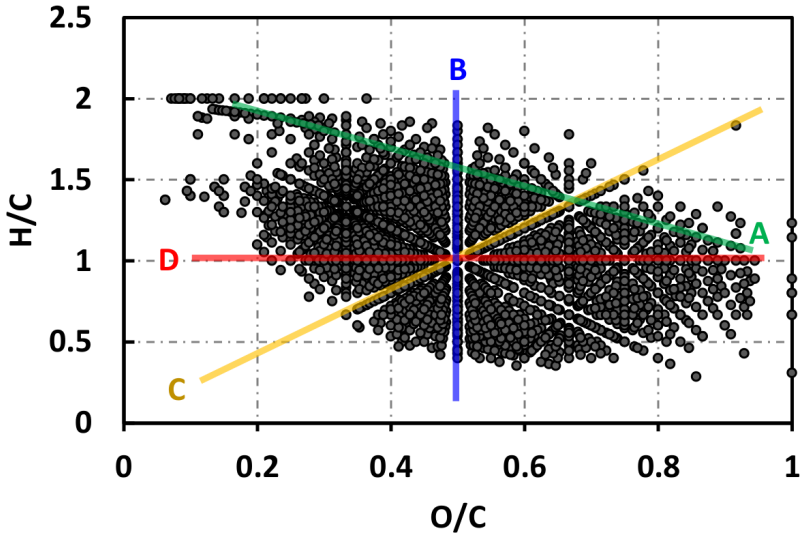


Figure 2.5: Van Krevelen plots of assigned molecular formulae in untreated SRFA (obtained with negative ion ESI FT-ICR mass spectrometry and MATLAB for CHO formulae assignment). Each circle represents a formula and the lines mark typical changes (reaction pathways) in the elemental ratios: A) methylation/demethylation; B) hydrogenation/dehydrogenation; C) condensation/hydrolysis; D) oxidation/reduction. Pathways from Sleighter et al., 2007 [95].

Furthermore, by having the size of each point (circle) representative of the relative intensity of the peak allows for some insight into the relative abundance of the molecular formulae. This is also useful when presenting data that includes heteroatoms. For example, Figure 2.6 depicts the different areas of the Van Krevelen space compounds including N and/or S encompass while also showing how species

including N and/or S have significantly smaller points, and thus contribute less to the total relative abundance of material compared to CHO only species.

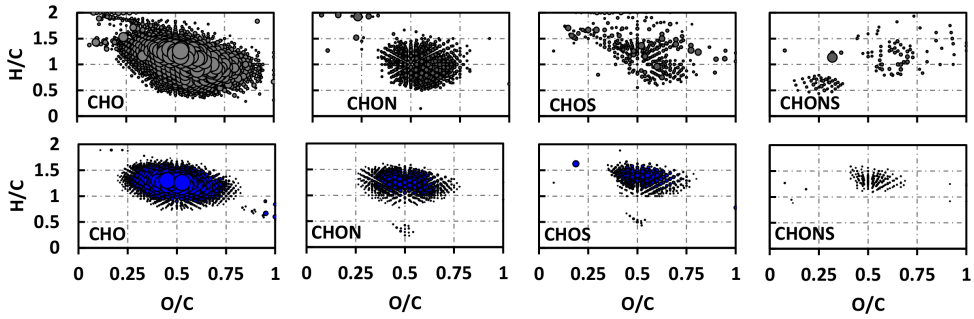


Figure 2.6: Van Krevelen plots of molecular formulae inclusions $C_{0-\infty}H_{0-\infty}O_{0-\infty}N_{0-1}S_{0-1}$ separated into four categories (obtained with negative ion ESI FT-ICR mass spectrometry and MATLAB for CHONS formulae assignment). **Top)** Untreated River (SRFA) shown in black; **Bottom)** Untreated Ocean (North Pacific Ocean depth of 1500 m) shown in blue.

DBE diagrams allow for a comparison of hydrogen saturation. For example, Figure 2.4 shows SRFA having DBE >20 while the ocean sample only has DBE <20, which shows that the ocean sample is overall more saturated than SRFA (a fully saturated compound would have DBE=0). The equation to calculate DBE is

$$DBE = c - h/2 + n/2 + 1 \tag{2.4a}$$

where c, h, and n are the number of carbon, hydrogen, and nitrogen atoms in a compound and DBE corresponds to the rings and/or double bonds in a compound [12]. All of these plots can be used to gather molecular information about DOM as well as help see the difference/similarities between such complex samples.

Chapter 3: Mass Labeling Method for Identifying Reduced Peaks and Molecular Formula Assignments (MATLAB Algorithms)

(Sections of this chapter are adapted from Bianca, M; Baluha, D; Gonsior, M; Schmitt-Kopplin, P; Del Vecchio, R; Blough, N.V. Contribution of ketone/ aldehyde-containing compounds to the composition and optical properties of Suwannee River fulvic acid revealed by ultrahigh resolution mass spectrometry and deuterium labeling. Anal. Bioanal. Chem. 2020, 412:1441–1451 [65].)

3.1 Reduction with Sodium Borodeuteride

Deuterium labeling was used to determine unique mass markers in mass spectra. Here, we describe a method which uses $NaBD_4$ to mass label compounds within DOM ensembles that contain borohydride-reducible functional groups such as ketones and aldehydes. As discussed in Chapter 1 section 1.3.3, selective chemical reduction has been shown to effect the optical properties and photochemical reactivity of DOM [6, 10, 57, 62, 64]; additionally, $NaBD_4$ has been previously shown to create unique mass markers, at $M + 3.0219n$ from the addition of nHD), in the ultrahigh resolution ESI mass spectra [64, 65]. These mass markers are unique in the mass spectra due to the negligible natural abundance of deuterium (unlike $NaBH_4$ which although reduces carbonyl containing species the same as

$NaBD_4$, does not create unique mass markers) [64].

These shifted masses arise directly from the reduction of ketone-and aldehyde-containing species by $NaBD_4$; deuterium adds to the carbonyl carbon and then the oxygen is protonated by the solvent (which in these reaction conditions is water) (Figure 3.1). These shifts are attributed to only the (irreversible) reduction of ketones and aldehydes, as hydroquinones should not affect the mass spectrum due to the rapid re-oxidation of quinones in the aerobic conditions (as well as deuterium exchange of the hydroquinones with solvent protons) [64].

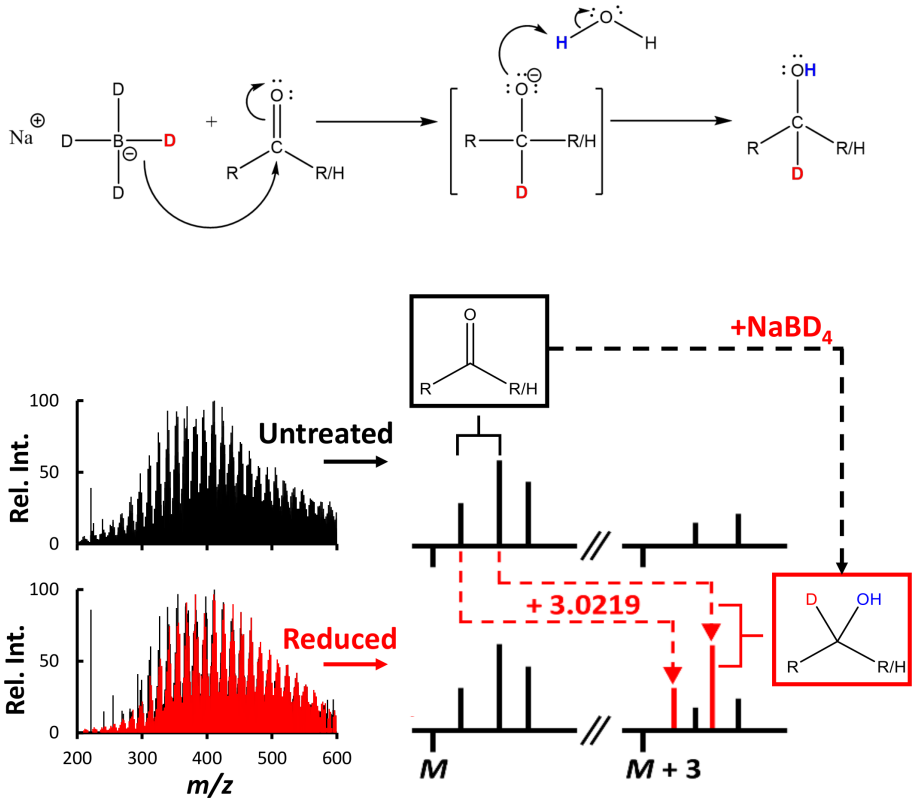


Figure 3.1: Top) Sodium borodeutride reduction scheme. **Bottom)** Left-Negative ion ESI FT-ICR mass spectra 200-600m/z of untreated SRFA sample (black) and reduced sample (reduced peaks in red, not reduced peaks in black). Right- Example of M+3 shifts due to the reduction of ketone and aldehyde containing species.

Reduction protocol. Specific reduction methods for each study are located in the methods section of each chapter. Briefly, the general protocol for carrying out reductions is as follows: Samples are prepared in Milli-Q water at $\sim$ 10 mg/L and filtered through a preconditioned 0.2 μm nylon filter. Optical measurements are taken pH$\sim$7 and pH$\sim$10 before dividing the sample into two aliquots; one to use for reduction and the other kept as an untreated control. Reduced samples are treated with solid $NaBD_4$ by adding the borodeuteride directly to the aqueous samples. Samples are then stored open to air in the dark for 24 hours (at room temperature) to allow for reduction (the untreated sample is kept in the same conditions for a control). After 24 hours, the optical measurements taken again and then the samples are adjusted to pH 2, which removes any residual $NaBD_4$. Then to remove excess borate salts, the samples are purified by solid phase extraction (SPE) [81, 96]. The absorbed material is eluted with methanol and split into two aliquots, one for MS and the other is dried down using gentle N_2 stream then redissolved in purified water for optical measurements.

3.2 Mass Spectrometric Data Acquisition and Pre-processing

The following mass spectrometric data acquisition and pre-processing was used for all samples in the following chapters. Ultrahigh resolution mass spectra of samples were obtained using a Bruker Solarix 12 Tesla Fourier transform ion cyclotron resonance mass spectrometer (FT-ICR MS) at the Helmholtz Center for Environmental Health (Munich, Germany), operating at a nominal resolution of >500,000

at m/z 400. An Apollo II electrospray ionization source was operated in negative ion mode using a spray voltage of -3.6 kV and sample was introduced to the electrospray at a flow rate of 120 μL min^{-1}. Mass spectra were recorded from 147–800 m/z and 500 individual scans were averaged to achieve the precision needed for mass assignments. All mass spectra were pre-calibrated using arginine clusters and the averaged spectra were again post-calibrated using a known list of always present DOM m/z ions (in the 200-800 m/z range-for calibrant list see Table A.1) to increase the mass accuracy which was always better than 0.2 ppm. All m/z peak lists were generated and post-calibrated using the Bruker Daltonics Data Analysis 4.0 software.

3.3 Molecular Formula Assignments and Identification of Reducible Species using MATLAB Algorithms.

The majority of this section has been published in reference [65].

All data processing and analyses were undertaken using custom-written MATLAB functions [65] (illustrated in Figure 3.3 below; for code see Appendix H), and described as follows:

Excision of multiply-charged *m/z* ions and common contaminants (function F1_list): Doubly-charged species were identified and excluded by searching for pairs of peaks in the raw samples (raw_UNT, raw_RED) which differed by 0.501678 m/z, respectively (i.e., the expected mass difference between doubly-charged species and their ^{13}C-containing isotopologues), within a

0.4 ppm error window. This was done to simplify the mass spectra, since the majority of the ions should be singly charged ($>\sim 90\%$). Also excluded from the peak lists were those that resided within 0.4 ppm of the major *m/z* ions present in the mass spectrum of the extraction blank (blk_major) (Table A.2 and Figure A.1).

Molecular formula assignment (function F2_form): Molecular formulae assignment of *m/z* values was accomplished using a "low-mass moiety" approach similar to that described by Perdue and Green [97, 98]. In this algorithm, each *m/z* was converted to its neutral mass by the addition of the mass of H+ (1.0072764 amu). From this neutral mass, a base molecular formula containing ^{1}H and ^{12}C with the lowest possible H/C molar ratio (e.g. $C_{28}H_6$ for a measured *m/z* of 341.0878) as well as all allowed combinations of ^{14}N, ^{2}H, and ^{34}S (where the maximum numbers of ^{14}N, ^{34}S, and ^{2}D, were pre-determined). Integer numbers of $+ O_3 - C_4$ and/or $+CH_4 - O$ modifications were made to this base molecular formula until the following requirements were satisfied: (1) calculated mass of the formula was within 0.4 ppm of the measured mass, (2) $c > 0$, (3) $2 \leq h \leq 2c+2$, (4) $o \leq c$, and (5) $o + n + s > 0$, where c, h, o, n, and s are the numbers of ^{12}C, ^{1}H, ^{16}O, ^{14}N, and ^{32}S, respectively. If more than one molecular formula was possible for a single *m/z* value, the formula with the least number of heteroatoms (N + S) was used. Afterwards, isotopologues containing a single ^{13}C atom were identified by finding *m/z* ions that were 1.003355 *m/z* (i.e., the mass of ^{13}C minus that of ^{12}C) higher than a *m/z* which was assigned

to a molecular formula. Similarly, ^{34}S-containing isotopologues were identified by searching for m/z ions that were 1.995980 m/z greater than a m/z value that was assigned to a ^{32}S-containing formula. Furthermore, including deuterium (D) in the possible combinations of atoms by allowing up to two HD subunits to be added into the base hydrocarbon molecular formula allowed reduced peaks to be labeled as not reduced, singly reduced, and/or doubly reduced. In preliminary tests (data not shown), this low-mass moiety algorithm yielded formulas identical to a "brute-force" algorithm which cycled through all the chemically reasonable combinations of C, H, O, N, and S, with the benefit of completing molecular formula assignment of a single spectrum in a few minutes rather than in several hours on a personal computer.

Identification of deuterium-containing reduced ketones/aldehydes (function F3_search and F4_search): For each m/z in UNT to which a molecular formula was assigned (MF_UNT), the mass spectrum of a borodeuteride-reduced sample (peaks_RED) was searched for m/z values at M+3.021927 and M+6.043854 within a 0.4 ppm window, corresponding to the expected masses of singly- and doubly-reduced analogues of the species at m/z M. If only the peak at M+3.021927 or M+6.043854 was found in the borodeuteride-reduced spectrum, then the peak at M in UNT was identified as comprising at least one singly-reducible or at least one doubly-reducible species, respectively (MFR_RED). If both corresponding ions were found, then the original ion at M was identified as comprising either a combination of species with one and two reducible moi-

eties, or an individual (or set of isobaric) species with two reducible moieties that was only partially reduced (example Appendix B Figure B.1). The *m/z* ions in the reduced (searched) mass spectrum that resided at M+3.021927 and M+6.043854 were assigned to the same formulae as that for the *m/z* ion at M in UNT with the addition of HD and H_2D_2, respectively. As a control, the peak list of UNT, rather than that of a reduced sample, was searched in a similar manner for the M+3.021927 and M+6.043854 mass markers.

The respective intensities of the reduced peaks were found (F4_search). Each peak identified as either a singly- or doubly-reduced species in the searched mass spectral peak list was given one of four mutually exclusive designations: 1O, 1B, 2O, or 2B. If only a peak at *m/z* of either M+3.021927 or M+6.043854 was found in the searched mass spectrum (for a given M), then that peak was designated as either 1O or 2O, respectively. However, if peaks at both M+3.021927 and M+6.043854 were found, then they were designated as 1B and 2B, respectively (Figure 3.2). The extent of reduction that occurred in each sample was then quantified by the percent number of *m/z* peaks (%N_X) and the intensity-weighted percent of total ions (%I_X) that were identified as either type 1O, 1B, 2O, and 2B, as follows:

$$\%N_X = 100 \times \frac{N_X}{N_{total}} \tag{3.1a}$$

$$\%I_X = 100 \times \frac{I_X}{TPI} \tag{3.1b}$$

where N_X and I_X are the number of peaks and the sum of peak intensities,

respectively, of the peaks identified as either 1O, 1B, 2O, or 2B, and N_{total} and TPI are the total number of peaks and total peak intensity, respectively, of the searched(reduced) mass spectral peak list (to which formulae were assigned).

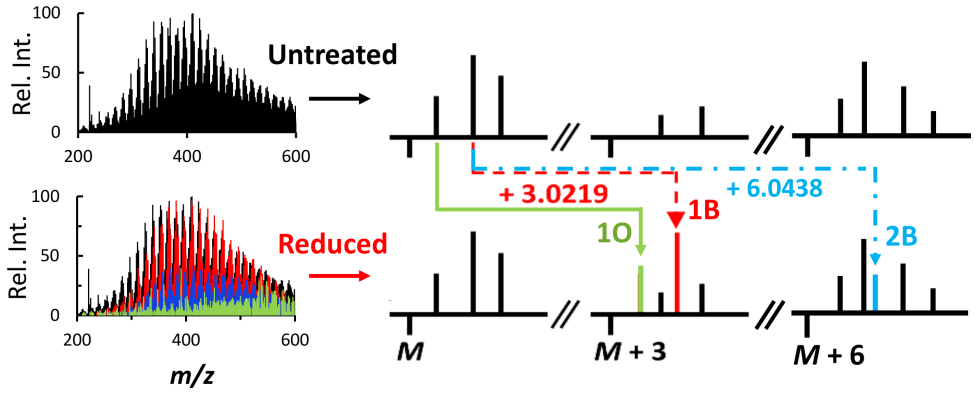
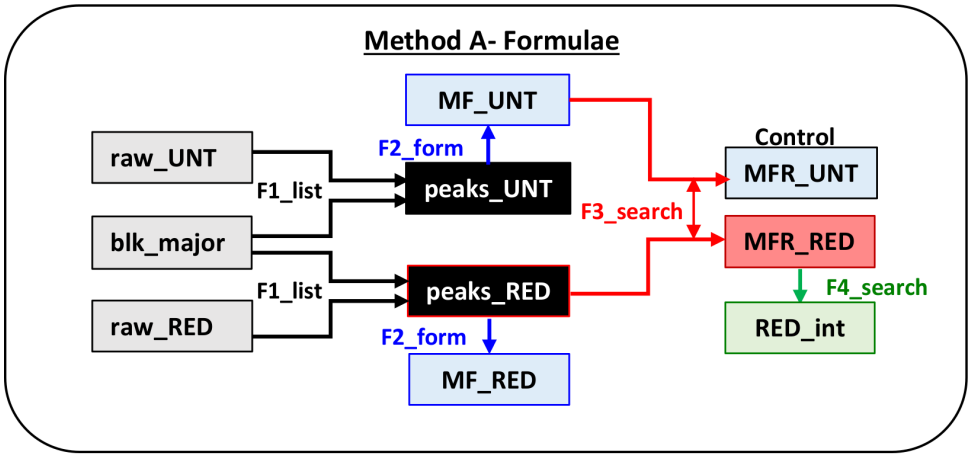


Figure 3.2: Mass labeling with sodium borodeuteride exclusive designations (1O,1B, 2B). **Left)** Negative ion ESI FT-ICR mass spectra 200-600 m/z of untreated SRFA sample (black) and reduced sample (reduced peaks in green, red, and blue, not reduced peaks in black). **Right)** Example of $M+3n$ shifts due to the reduction of ketone/aldehyde containing species: 1O (green), 1B (red), and 2B (blue). 2O not shown (less than 2% of the peak intensity for all samples).

The internal consistency of this method was checked using two additional methods (see Appendix B.1 section C.1 Table C.1 and Figure C.1) The above molecular formula assignment and mass labeling method (Method A) is used for all subsequent analysis in the following chapters.



raw_X: raw mass spectral data of sample X (*m/z* and intensity of ions with S/N>10)
blk_major: major peaks in the blank
peaks_X: blank-subtracted list of singly-charged ions in raw_X
MF_UNT: molecular formula of ions in peaks_UNT, calculated by low-mass moiety algorithm
MF_RED: molecular formula of ions in peaks_RED, calculated by low-mass moiety algorithm
MFR_UNT: molecular formula of ions in MF_UNT, labeled as non, singly, and/or doubly reduced upon treatment by borodeuteride
MFR_RED: molecular formula of ions labeled as non, singly, and/or doubly reduced upon treatment by borodeuteride in reduced sample
RED_int: molecular formula of ions labeled in MFR_RED with the intensity of the respective reduced species.

Figure 3.3: Illustration of data variables (boxes) and functions (arrows) used for the analysis of mass spectrometric data (performed in MATLAB) for untreated sample (UNT) and reduced sample (RED).

Chapter 4: Examining the contribution of ketone- and aldehyde-containing species to the composition and optical properties of Suwannee River fulvic acid as revealed by revealed by ultrahigh resolution mass spectrometry and deuterium labeling

(Adapted from Bianca, M; Baluha, D; Gonsior, M; Schmitt-Kopplin, P; Del Vecchio, R; Blough, N.V. Contribution of ketone/ aldehyde-containing compounds to the composition and optical properties of Suwannee River fulvic acid revealed by ultrahigh resolution mass spectrometry and deuterium labeling. Anal. Bioanal. Chem. 2020, 412:1441–1451 [65].)

4.1 Introduction

This Chapter implements the mass labeling method described in Chapter 3 section 3.3 for the detection of species containing borodeuteride-reducible functional groups (e.g. ketones and aldehydes) in Suwannee River fulvic acid (SRFA). SRFA was reduced using increasing concentrations of borodeuteride, ranging from 2-157 fold sodium borodeuteride ($NaBD_4$) mass excess/mass SRFA. The optical changes observed after reduction were monitored and negative mode electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry (ESI FT-ICR MS) was used for the detection of reduced species. The extent of reduction observed via

ESI FT-ICR MS with increasing amounts of $NaBD_4$ correlated well with changes in optical absorption and emission parameters, revealing new insights concerning the composition and contribution of ketone/aldehyde functional groups to the bulk optical properties of SRFA.

4.2 Materials and Methods

4.2.1 Reagents and Materials

SRFA, an extract from XAD-8 resin, was purchased from the International Humic Substances Society, (Standard II, catalog number: 2S101H). Perchloric acid ($HClO_4$), Ammonium hydroxide (NH_4OH), sodium borodeuteride ($NaBD_4$; 98% deuterium, 90% purity), and quinine sulfate were purchased from Sigma Aldrich. High purity methanol (MeOH) was purchased from Honeywell Burdick and Jackson. Purified water was obtained from a Milli-Q purification system (Millipore). Bond Elut PPL cartridges– a proprietary functionalized styrene-divinylbenzene polymer (100 mg, part number 12105004) were purchased from Agilent. Tumeric 110 Analytical Test Strips were purchased from Scientific Equipment of Houston (Lot 4012).

4.2.2 Sample Preparation of Suwannee River Fulvic Acid

A simple scheme of sample preparation is depicted below in Figure 4.1. SRFA solutions were prepared in Milli-Q water (10 mg/L) and filtered through a preconditioned 0.2 μm nylon filter. Optical measurements were taken at pH$\sim$7 and pH$\sim$10 before splitting the sample into aliquots (untreated and reduced). For the reduced

(Rx) samples, aliquots were treated with 2, 3, 11, 23 (duplicate), 50, 52, and 157- fold mass excesses of $NaBD_4$ relative to SRFA mass (R2x, R3x, R11x, R23xA, R23xB, R50x, R52x, and R157x respectively). Since the reduction of SRFA was carried out by adding solid borodeuteride directly to the aqueous samples, it was difficult to prepare exact replicates of the reduced samples. Therefore, R2x and R3x, R23xA and R23xB, and R50x and R52x are assumed to be as close to sample replicates as practically achievable. These samples were used to approximately assess the reproducibility of the methodology. To allow for reduction, $NaBD_4$ reduced samples were stored open to air in the dark for 24 hours and kept at room temperature as were the untreated (UNT) samples for a control. After 24 hours, optical measurements were taken at pH 10 and then again after adjusting the pH down to 7 with $HClO_4$ (1N). Afterwards, samples were adjusted to pH 2 (to remove any residual $NaBD_4$) with $HClO_4$ (1N). Each sample was then purified by solid phase extraction (SPE) to remove excess borate salts [81, 96]. Briefly, a 100 mg PPL cartridge was preconditioned with 5 mL of MeOH and 5 mL of 0.1% formic acid, and then 5 mL of the pH$\sim$2 sample was eluted by gravity. The cartridges were then rinsed with an additional 10 mL water with 0.1% formic acid, dried with a gentle nitrogen stream (15-20 min). The absorbed organic matter was eluted with 2 mL methanol (the first 1 mL was collected and split for the optics and MS). Half (0.5mL) of the DOM methanol eluate was stored (-18°) for FTICR MS analysis, while the other half (0.5 mL) was dried using a gentle N_2 stream then dissolved in purified water (2.5 mL) for optical measurements (pH$\sim$7 and pH$\sim$10). Absence of residual borate was confirmed by using turmeric test strips.

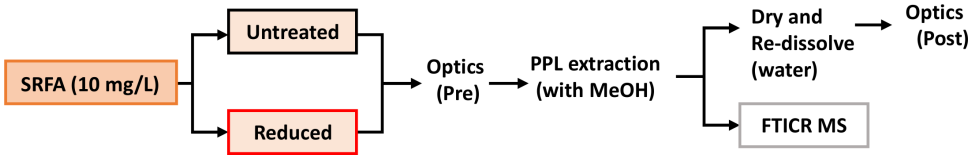


Figure 4.1: Block diagram of the sample preparation of untreated and reduced SRFA.

4.2.3 Mass Spectrometric Data Acquisition, Pre-processing, Molecular Formula Assignments, and Identification of Reducible Species in SRFA

Mass spectrometric data acquisition and pre-processing for SRFA. Mass spectrometric data acquisition and pre-processing details for untreated (UNT) and borodeuteride-reduced (Rx, 2-157 fold mass $NaBD_4$/mass SRFA) samples are provided in Chapter 3 section 3.2.

Molecular formula assignments and identification of reducible species in SRFA. Data processing and analyses for molecular formula assignments and identifying ketone/aldehyde reducible species were undertaken using custom-written MATLAB functions described in detail in Chapter 3 section 3.3 (MATLAB code provided in Appendix H).

4.2.4 Optical Measurements for SRFA

Apparatus for optical measurements. For instruments employed for absorbance and fluorescence measurements see Appendix C.1 section D.1.

Optical measurements. Parameters and protocols used for both absorbance and fluorescence measurements are detailed in Appendix C.1 section D.2.

Comparing pre- and post-extraction material. Optical measurements were taken on samples prior to (Pre) and following (Post) SPE whereas all mass spectra were collected exclusively on the Post-SPE samples (shown in above in Figure 4.1). As mentioned above, SRFA is an extract from XAD-8 resin and while the SPE procedure used in this experiment employed PPL cartridges, which are functionalized polystyrene divinylbenzene (not XAD-8 resin), only minor effects were expected to be seen due to SRFA already being an extract.

Preliminary experiments showed that SPE had relatively minor effects on the measured optical properties of SRFA (Appendix E Figure E.1). Absorption of extracted material (post) was normalized to the absorption at 300 nm of its respective pre-SPE counterpart due to dilution and therefore, the normalized absorbances are used for analysis (further details provided in Appendix E section E.0.1).

4.3 Results and Discussion

4.3.1 General Mass Spectrometric Features

The mass spectra of all the samples contained several thousand m/z ions from 150-800 m/z, but only the 200-600 m/z range was analyzed. This range was examined because >85% of the total peak intensity (TPI) (Table 5.1) was found within this range (excluding R157x), and at $m/z > 600$ (at a 1ppm error), there is a far greater possibility of acquiring more than one molecular formula assignment per

m/z value (the assignment becomes equivocal) [91,99]. It is apparent that the average m/z and TPI for reduced samples were generally substantially lower than that of the untreated samples. These two observations are somewhat counterintuitive, since reduction of a ketone/aldehyde functional group by borodeuteride will result in a 3-mass unit increase in molecular weight as well as the fact that generally, an alcohol can be expected to be more efficiently ionized via ESI than an analogous carbonyl. However, it is most likely that most carbonyl-containing species also contain carboxylic acid functional groups as well, and that ionization is primarily due to deprotonation of those acidic groups. The overall distribution of m/z peaks in our mass spectra are similar to those of previous studies using ESI-MS [73, 79]. However, it should be noted here that the m/z distributions observed by ESI MS can vary significantly not only with the hardware configuration of the instrument but also many experimental parameters (e.g. cone voltage [100] (therefore, comparisons between studies are approximate).

Table 4.1: Mass spectral peak distributions.

Sample	Peaks (150 – 800 *m/z*)				Peaks (200-600 *m/z*)[a]			
	M_N[b]	M_W[c]	M_N/M_W	TPI[d]	M_N[b]	M_W[c]	M_N/M_W	TPI[d]
UNT	465	493	0.94	1.5E+11	431	448	0.96	1.4E+11 (86%)
R-2x	433	451	0.96	3.6E+10	425	440	0.97	3.5E+10 (96%)
R-3x	458	478	0.96	1.6E+11	444	459	0.97	1.5E+11 (93%)
R-11x	464	488	0.95	9.6E+10	440	456	0.96	8.5E+10 (88%)
R-23Ax	428	448	0.96	3.3E+10	421	436	0.97	3.2E+10 (96%)
R-23Bx	419	436	0.96	1.9E+10	415	429	0.97	2.1E+10 (98%)
R-50x	449	468	0.96	1.5E+11	438	452	0.97	1.5E+11 (94%)
R-52x	456	474	0.96	1.2E+11	443	458	0.97	1.1E+11 (94%)
R-157x	407	423	0.96	1.9E+10	404	417	0.97	1.7E+10 (98%)

[a]singly-charged, blank subtracted, error < 0.4 ppm; [b] M_N: number-averaged *m/z*; [c] M_W: weight-averaged *m/z*; [d] TPI: total peak intensity; TPI of used peaks: includes percent of TPI for all peaks in parenthesis

4.3.2 Assessment of Molecular Formulae Algorithm Assignments

Prior to molecular formula assignments, the maximum allowed number(s) of non-oxygen heteroatoms (N and S) as well as number of possible deuterium atoms is specified. While allowing large numbers of heteroatoms can be beneficial for assigning formulae to as many molecular ions in the mass spectrum as possible, doing so may result in many false and/or equivocal assignments, especially at high *m/z*

($m/z > 600$) [91, 99]. Therefore, the molecular formula assignment algorithm was run using different maximum allowances of ^{14}N, ^{32}S, and D. R50x was chosen as a representative of the reduced samples for this preliminary study. When the formula algorithm was constrained to give only CHO-containing formula and up to two D atoms were allowed, the %TPI of UNT and R50x that was assigned D-containing formula was <1% and 48%, respectively. The allowance of up to three and four D-atoms per molecule resulted in minimal increases in the %TPI assigned molecular formula in the reduced sample (<3%), but produced also an increase in more equivocal assignments as well as a slight increase in (false) identifications of D-containing species in UNT (Figure 4.2). These results suggest that although this analysis is less definitive if more than two D-atoms is allowed, it is unlikely that a significant number of ionizable DOM species in the analyzed mass range contain more than two borodeuteride-reducible functional groups. When additional heteroatoms (formulae containing up to 2 N or 1 S) were included, they accounted for less than a 9% increase of the TPI assigned, while formulas with only CHO accounted for the majority (>85%) of the TPI in the 200-600 m/z range. Furthermore, species containing only CHO made up over 99% of the TPI for the assigned reduced molecular formulae (Figure 4.2). Therefore, CHO only species and D up to 2 in the 200 – 600 m/z range (Appendix E Figure E.2) were analyzed for subsequent analysis.

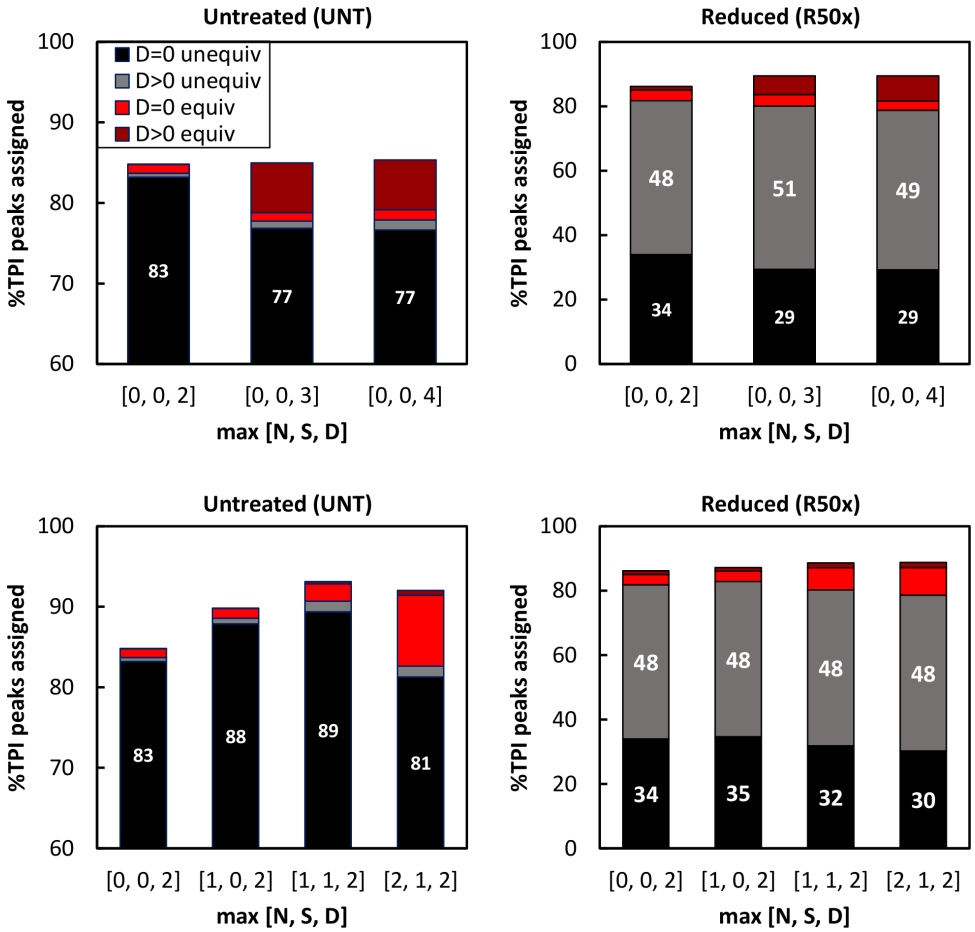


Figure 4.2: Bar graph of %TPI of UNT (left) and R-50x (right) assigned equivocal/unequivocal and D-containing / non-D-containing molecular formulas for various allowances of N, S, D. **Top)** Different allowances of deuterium (2-4). **Bottom)** Different allowances of heteroatoms (N,S) and constant D (2): N, S, D (0,0,2; 1,0,2; 1,1,2; 2,1,2) from 200-600 m/z.

4.3.3 Reduction and Identification of Ketone/aldehyde-containing Species in SRFA.

As discussed in Chapter 3, selective chemical reduction of ketone/aldehyde-containing species using $NaBD_4$ has been shown previously to create unique mass

labels for species that undergo reduction of one and/or two ketone/aldehyde moieties [64]. Species with borodeuteride-reducible moieties were identified, as described in Chapter 3 section 3.3, by searching the m/z (singly-charged) ions of the reduced samples for molecular ions at M+3.021927 and M+6.043854 (within 0.4 ppm) (Figure 4.3 and Figure 4.4). The untreated sample was used as a control by searching for these unique mass markers using the untreated m/z list. The control (UNT) search falsely identified less than 1% of CHO molecular formulae as reduced species (Figure 4.2, Untreated [0,0,2]) and therefore resulted in a very low $\%N_{total}$ (sum of $\%N_{1O}$, $\%N_{2O}$, $\%N_{1B}$, and $\%N_{2B}$) and $\%I_{total}$ (sum of $\%I_{1O}$, $\%I_{1B}$, $\%I_{2O}$, and $\%I_{2B}$) indicating that this method generates very few false positives.

At low mass excess of $NaBD_4$ (R2x), the majority (∼99%) of the reduced peaks are observed at M+3 (1O) (Figure 4.4). At higher mass excesses of $NaBD_4$, both M+3 and M+6 peaks were observed (1B and 2B respectively). As the amount of $NaBD_4$ increases, the intensity of the M+6 peaks (2B) increased, while the intensity of the M+3 only (1O) peaks decreased until ∼50-fold mass excess, remaining approximately constant thereafter (Figure 4.4, Figure 4.5, and Figure 4.6). These combined results suggest that some species were reduced once under low amounts of $NaBD_4$ and twice under higher amounts of $NaBD_4$ (shown by the arrows in Figure 4.3), due to the subsequent reduction of additional, less reactive ketone/aldehyde moieties within the same species.

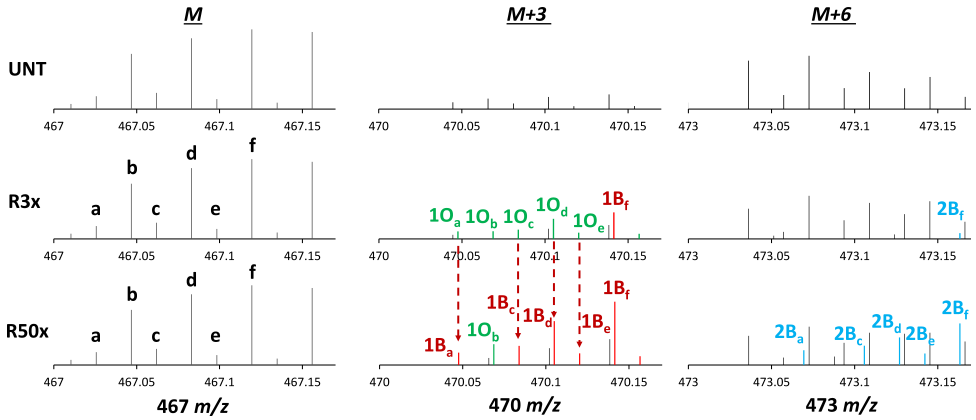


Figure 4.3: Mass spectra (expanded region 467-473 m/z) of UNT and reduced (R3x and R50x) SRFA at 467 (left; M), 470 (middle;M+3), and 473 (right; M+6) m/z. Black peaks represent untreated peaks in UNT and not-reduced peaks in reduced samples; peak intensity of singly-reduced peaks shown by **1O-green** and **1B-red**. Red arrows depict the change from an M+3 only peak (1O) under low $NaBD_4$ (R3x) to an M+3 with an M+6 peak (1B) under high $NaBD_4$ (R50x). Intensity of the respective doubly reduced peaks shown by **2B-blue**.

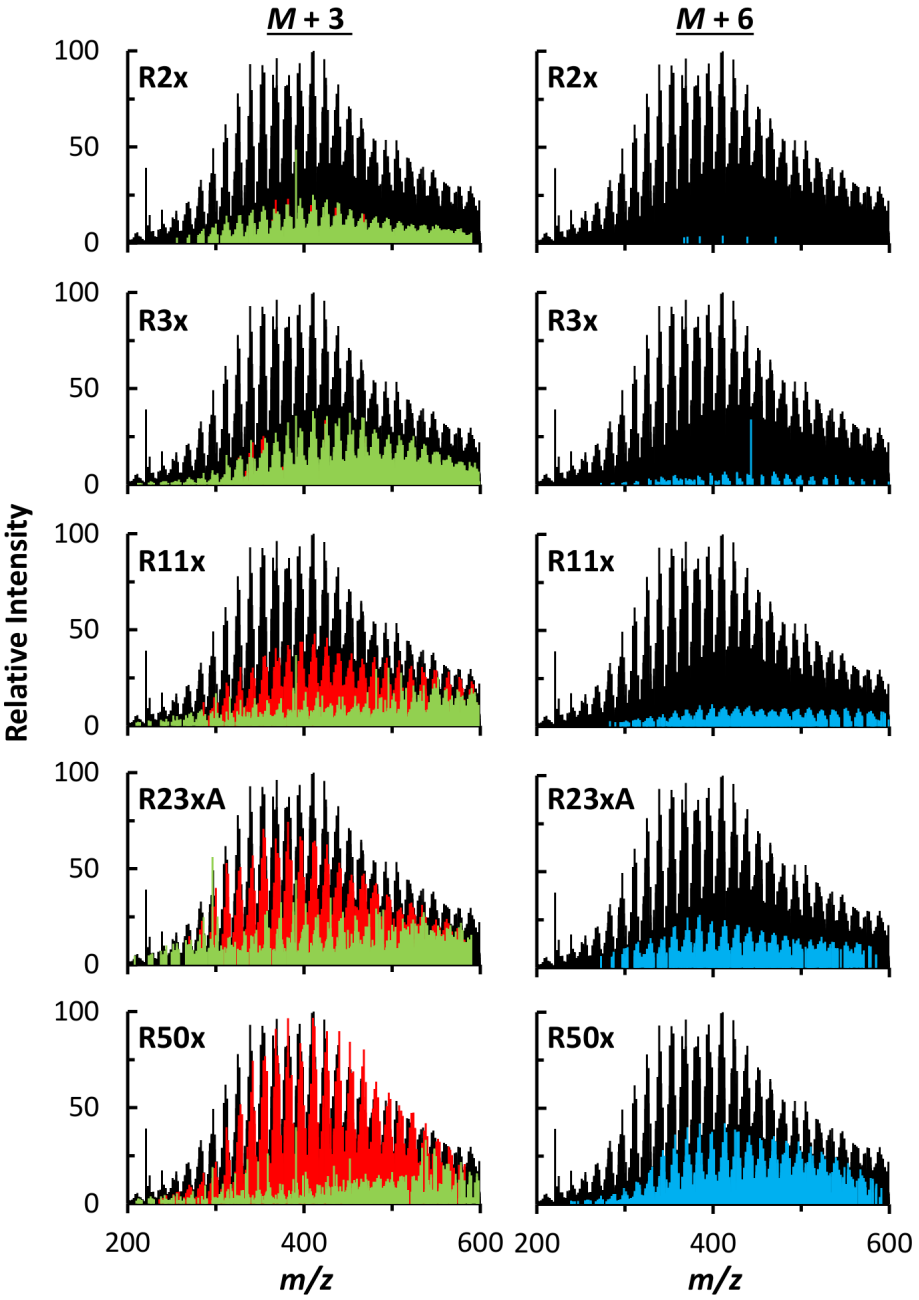


Figure 4.4: Negative ion ESI FT-ICR mass spectra (200-600 m/z) of peaks assigned CHO molecular formulae. Untreated sample (black), overlaid with samples reduced (Rx) for 24 hours (increasing from 2- to 50-fold $NaBD_4$ mass excess/mass SRFA). Intensity of reduced peaks: 1O (green, M+3 only), 1B (red, both M+3 and M+6, showing intensity of M+3 peak), and 2B (blue, both M+3 and M+6, showing intensity of M+6 peak). 2O not shown (less than 1% of the peak intensity for all samples).

Furthermore, the constant values observed after ∼50-fold mass excess for groups, %I1O and %N1O, indicates some species (approximately 12%, Figure 4.5 and Figure 4.6) only had one reducible group or only one "accessible" reducible group (1O). Species that were doubly-reduced only (2O) did not significantly contribute to %I2O and %N2O (<1%) (Figure 4.5 and Figure 4.6). Most peaks had both M+3 and M+6 peaks (1B and 2B, respectively). These peaks (1B and 2B) increased in intensity (%I1B and %I2B) up to ∼50-fold mass excess; and this result suggests most peaks comprised either a combination of species with one and two reducible moieties, or an individual (or set of isobaric) species with two reducible moieties that were only partially reduced (as mentioned in Chapter 3 section 3.3). This result also indicates that the reduction is incomplete and species with two reducible groups remain unreduced unless high amounts of $NaBD_4$ are added. Possibly, treatment of SRFA, and perhaps other DOM samples, with relatively low mass equivalents of $NaBD_4$ can be used to identify the more reactive ketone/aldehyde-containing species, whereas reduction with higher amounts $NaBD_4$ can be used to differentiate among species with one and/or two reducible (ketone/aldehyde) functional groups.

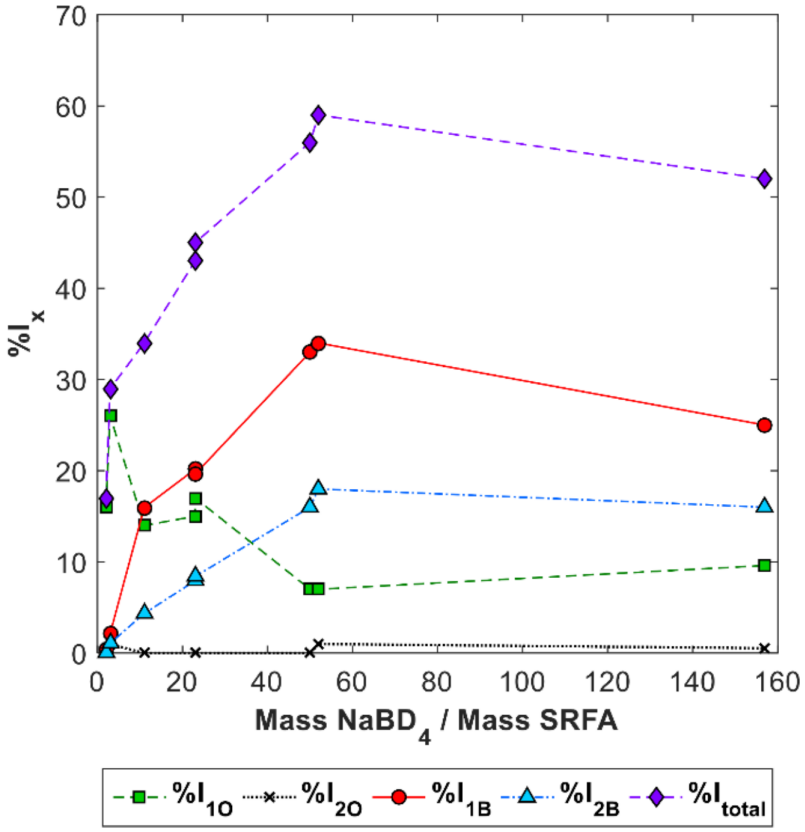


Figure 4.5: $\%I_X$ extent of reduction vs increasing mass excess $NaBD_4$ (2 to 157-fold) added to reduced samples. $\%I_X$ (intensity-weighted percent of ions across 200-600 m/z). Total percent of reduced intensity (purple, $\%I_{total}$) allocated into four groups: $\%I_{1O}$ (green) % of intensity of M+3 only peaks), $\%I_{2O}$ (black) % of intensity of M+6 only peaks; $\%I_{1B}$ (red) both M+3 and M+6 found, showing % of intensity of M+3 peaks, $\%I_{2B}$ (blue) both M+3 and M+6 found, showing % of intensity of M+6 peaks.

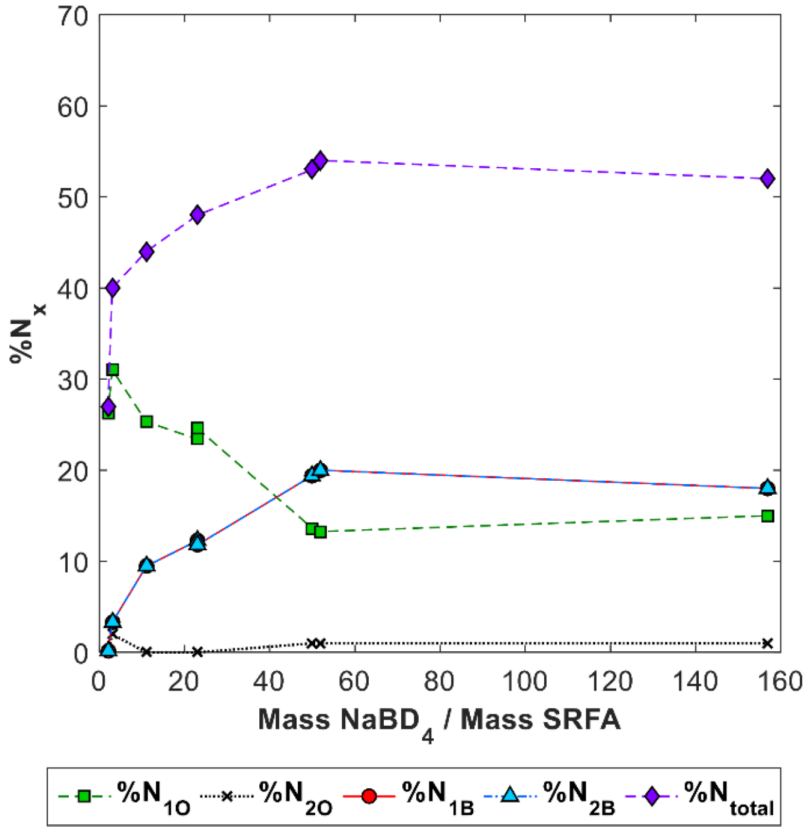


Figure 4.6: $\%N_X$ extent of reduction vs increasing mass excess $NaBD_4$ (2 to 157-fold) added to reduced samples. $\%N_X$ (fraction of the number of reduced ions across 200-600 m/z). Total percent of the number of reduced ions (purple, $\%N_{total}$) allocated into four groups: $\%N_{1O}$ (green) % of intensity of M+3 only peaks), $\%N_{2O}$ (black) % of intensity of M+6 only peaks; $\%N_{1B}$ (red) both M+3 and M+6 found, showing % of intensity of M+3 peaks, $\%N_{2B}$ (blue) both M+3 and M+6 found, showing % of intensity of M+6 peaks.

4.3.4 Molecular Composition of Ketone/aldehyde-containing Species in SRFA

Van Krevelen plots (i.e. plots of H/C versus O/C molar ratios) and double bond equivalent (DBE) plots were constructed for R3x, R11x, and R50x, to visualize

the composition and the increase of peaks arising from reduction, with the size of the circle relative to the peak height (Figure 4.7). The H/C and O/C of the majority of all formulae were centered around H/C = 1 and O/C = 0.5, which has been observed previously for SRFA and other terrestrially influenced DOM samples [49, 99].

Reduced species with the highest intensity peaks (largest circles), were near the center of the Van Krevelen plots, indicating that many of these species are easily reducible (even under low mass equivalents of $NaBD_4$). For all reduced samples, a greater degree of reduction was observed to occur at the higher masses (Figure 4.7 400-600 m/z) due to these species having more reactive centers that can be reduced.

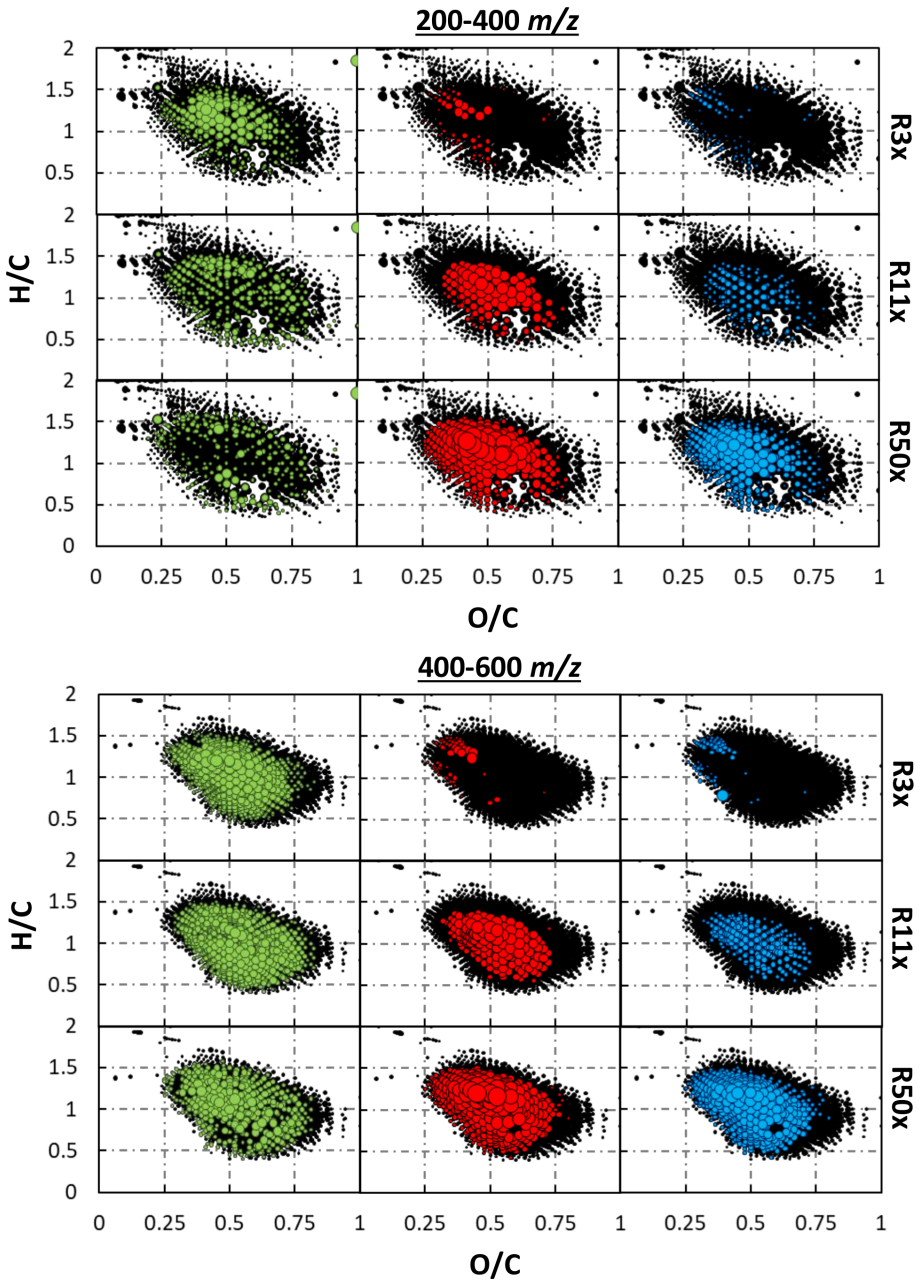


Figure 4.7: Top) Van Krevelen plots of CHO formulae identified in low (200-400 m/z) and **Bottom)** high (400-600 m/z) mass ranges. Untreated sample shown in black, overlaid with the intensity of the respective 24 hour-reduced peaks arising from reduction by borodeuteride (R3x, R11x, and R50x). Intensity of reduced peaks color coded: 1O (green, M+3 only),1B (red, both M+3 and M+6, showing intensity of M+3 peak), and 2B (blue, both M+3 and M+6, showing intensity of M+6 peak). 2O not shown (less than 1% of the peak intensity for all samples).

A similar distribution of peaks was observed for M+3 only species (1O) for all reduced samples, though in the 200-400 m/z range there is a significant decrease in the number of peaks with increasing $NaBD_4$ (Figure 4.7 and 4.8) as seen above (Figure 4.4). In contrast, peaks that had both M+3 and M+6 had substantially higher numbers of reduced peaks as well as increased intensity of these peaks (1B and 2B, respectively) with increasing amounts of $NaBD_4$ (Figure 4.7). Furthermore, the distribution, across the entire mass range (200-600 m/z) of these peaks (1B and 2B) at low amounts of $NaBD_4$ (R3x) was above H/C$\sim$1 and below DBE$\sim$10 while with increasing amounts of $NaBD_4$ reduced peaks extended below H/C $\sim$1 and above DBE$\sim$ 10 indicating larger amounts of $NaBD_4$ act on different species (Figure 4.7 and Figure 4.8). The dramatic differences between the low and high $NaBD_4$ reductions, suggests that 1B and 2B species tend to contain multiple reducible moieties with varying reactivity and accessibility as a result of a high degree of structural diversity within them.

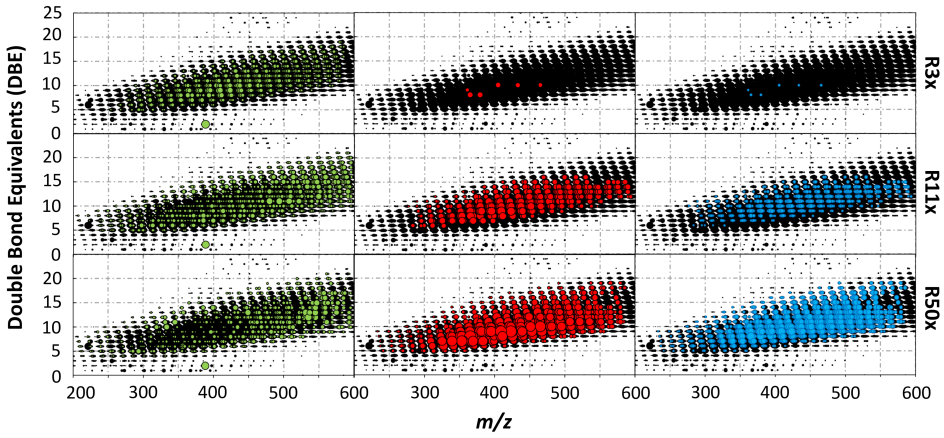


Figure 4.8: Double Bond Equivalents (DBE) vs m/z of CHO formulae identified. Untreated sample shown in black, overlaid with the intensity of the respective 24 hour-reduced peaks arising from reduction by borodeuteride (R3x, R11x, and R50x). Intensity of reduced peaks color coded: 1O (green, M+3 only),1B (red, both M+3 and M+6, showing intensity of M+3 peak), and 2B (blue, both M+3 and M+6, showing intensity of M+6 peak). 2O not shown (less than 1% of the peak intensity for all samples).

4.3.5 Relation of Identified Ketone/aldehyde-containing Species in SRFA to the Observed Optical Properties

Changes in bulk optical properties (absorption and emission) were investigated in relation to the structural information obtained by ESI FT-ICR MS. Borodeuteride reduction of SRFA resulted in a loss of absorption at wavelengths $>$ 250 nm and preferential loss at longer wavelengths (Figure 4.9) while the emission of the reduced samples blue shifted and increased in intensity with increasing amounts of borodeuteride, consistent with prior work [10, 40, 50, 57, 60, 61, 101] (Figure 4.10).

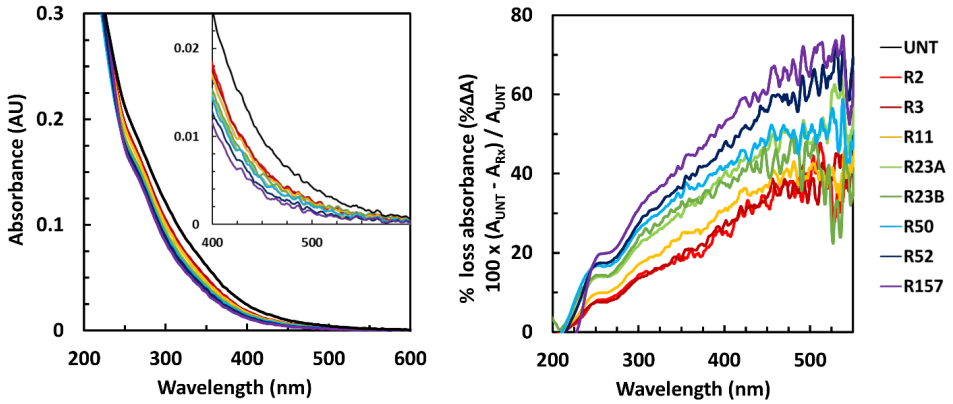


Figure 4.9: A) Absorption (prior to solid phase extraction-PPL) of untreated (UNT) and 24 hour-reduced (Rx) samples. **B)** Percent loss of absorption (relative to UNT) following reduction with increasing amounts of borodeuteride (2 to 157-fold $NaBD_4$ mass excess/mass SRFA respectively).

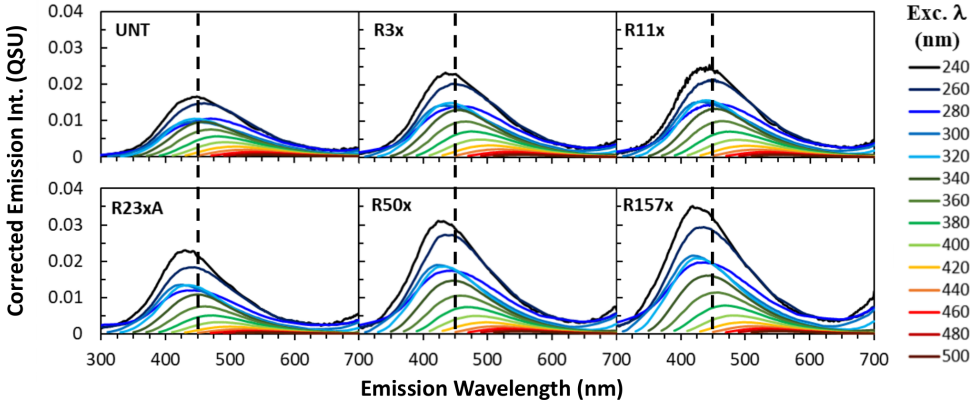


Figure 4.10: Corrected emission spectra (prior to solid phase extraction-PPL) of untreated (UNT) and reduced (Rx) SRFA.

There was a logarithmic dependence between the change in absorbance (%ΔA) and increasing $NaBD_4$ (Figure 4.11) that was similar to that previously reported for the changes in optical properties of several terrestrial and aquatic DOM samples as function of borohydride excess [50,101]. Similarly, $\%I_{total}$ and $\%N_{total}$ were observed

to have a logarithmic dependence with increasing $NaBD_4$ (Figure 4.11).

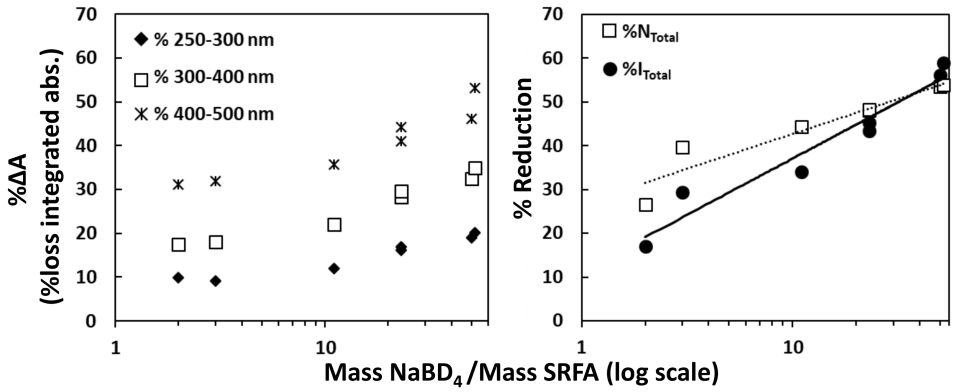


Figure 4.11: Logarithmic relationship between absorbance measurements as well as mass spectrometric measurements vs increasing amounts $NaBD_4$ (2 to 157-fold $NaBD_4$ mass excess/mass SRFA respectively) displayed on a log scale **Left)**. Percent loss of integrated absorbance (%ΔA) upon reduction by $NaBD_4$ relative to the untreated sample across three wavelength regions. **Right)** Intensity weighted percent of ions ($\%I_{total}$) (black triangles) and the fraction of the number of reduced ions ($\%N_{total}$) (red squares) across 200-600 m/z.

Although, extraction (SPE) showed minor effects on the measured optical properties of SRFA, $\%I_{total}$ was compared to both pre- and post-SPE optical measurements. $\%I_{total}$ correlated well (R^2>0.9) with several absorption and emission parameters (Appendix E Figure E.3). Furthermore, $\%I_{total}$ was observed to correlate well (R^2> ∼0.9) with the percent loss of absorption upon reduction within three wavelength regions (Figure 4.12). Separating $\%I_{total}$ into the four allocated groups (Appendix E Figure E.4), there is poorer correlation for %I1B and %I2B (R^2 ∼0.58-0.84), a negative correlation is obsereved for %I1O (singly-reduced only species) due to the loss of 1O species with increaseing amounds of $NaBD_4$, and no correlation for %I2O (doubly-reduced only species) as they only account for less than 1% of

the peak intensity for all samples. This result suggests that the optical properties are not significantly influenced by one group specifically, though they seem to be impacted more by species that have both M+3 and M +6 reducible peaks.

Strong correlations were observed between $\%I_{total}$ and the optical measurements, supporting the conclusion that reduction of ketone/aldehyde-containing species within SRFA are primarily responsible for the observed changes in absorption and emission as proposed in our previously described model (most likely via participation in chromophore-chromophore interactions) [40, 60].

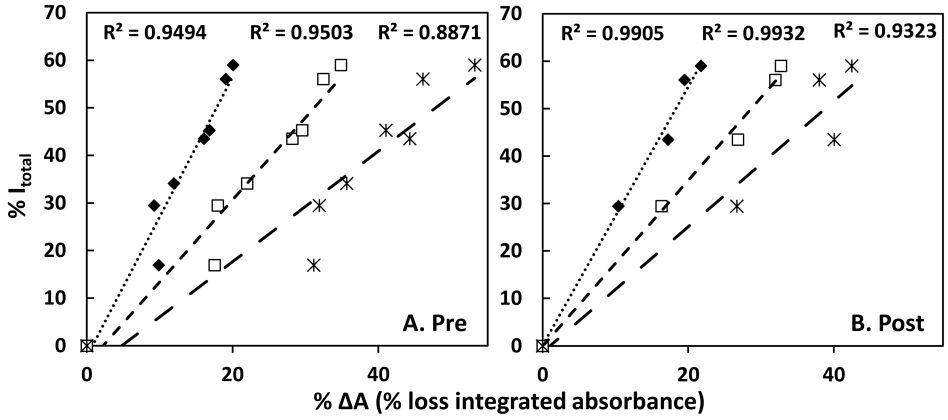


Figure 4.12: Percent of integrated absorbance loss upon reduction with $NaBD_4$ relative to untreated (UNT) sample (pH7) across three wavelength ranges correlated to $\%I_{total}$. 250-300 nm (black diamonds), 300-400 nm (white squares), and 400-500 nm (black asterisks) **A)** Pre: absorbance prior to solid-phase (PPL) extraction (excluding R157) and **B)** Post: absorbance after solid-phase (PPL) (data only collected for UNT, R3x, R23xA, R50x, and R52x).

4.4 Conclusions

The mass labeling method was successfully applied to gain detailed structural information of components within SRFA. The assessment of the molecular formula identified as being reduced revealed that the majority of assigned reduced compounds consist of CHO only species. In addition, reducing aqueous samples of SRFA with increasing concentrations of $NaBD_4$ gave insights into differences in both the presence and reactivity of ketone-/aldehyde-containing species. Van Krevelen diagrams were utilized for the visualization of these differences in the ketone-/aldehyde-containing species (from low to high $NaBD_4$) whose molecular formulae were shown to lie in different regions of an H/C versus O/C molar ratio plot. Furthermore, we found good correlations between the information obtained by this mass spectrometric method to bulk absorption data, especially between $\%I_{total}$ and the percent change in abosorbance, for both pre and post samples.

Chapter 5: Effects of Irradiation and Combined Irradiation *and* Reduction on the Optical and Mass Spectral Properties of SRFA

5.1 Introduction

As discussed in Chapter 1, DOM is a major pool within the global carbon cycle and is known to affect many biogeochemical processes [19,30–32]. The light absorbing fraction, chromophoric dissolved organic matter (CDOM), plays an essential role in many physical and chemical processes that greatly influence aquatic environments [30,32,36,102]. Photochemical reactions produce reactive oxygen species (e.g., singlet dioxygen, hydroxyl radicals, superoxide radicals) [58, 66, 103–105]. Studies have provided evidence that these reactive oxygen species produced by CDOM are involved in the degradation of the terrestrially-derived CDOM/DOM through the transformation of its chemical composition [106]. CDOM undergoes photodegradation [16, 107, 108] which converts high molecular weight DOM into lower molecular weight compounds [40, 103] as well as inducing photobleaching, resulting in the loss of absorption across the UV and visible regions [103]. Studies have suggested that photodegradation is likely to be an important removal route or modification pathway

for terrestrially-derived CDOM in the aquatic environment [37, 106, 109, 110]. Reports have stated that the photodegradation products (of terrestrial material) can be consumed by bacteria [111, 112] leading to further microbial processing; while others state that the irradiation of aquatic material appears to inhibit microbial processing [108, 113]. Recent studies have shown that photochemical transformations modify terrestrial DOM so that it better resembles marine DOM after irradiation [106] and that photochemical modifications of terrestrial DOM material resemble compounds found in marine environments [106, 114]. Currently, the effects of UV irradiation on the structure, composition and size distribution of DOM in the marine environment remains incomplete. [103, 115].

To better understand the processes leading to photodegradation and the overall cycling of CDOM, it is necessary to be able to detect and characterize the species within DOM that are altered photochemically, which in turn can potentially be used to gain insight into the structural basis of the optical properties of CDOM from the open ocean. This study uses the combination of bulk optical properties and ultrahigh resolution mass spectrometric analyses, as previously reported in Chapter 4 [65]. Here, how irradiation affects the optical and mass spectrometric properties of Suwannee River fulvic acid (SRFA) is examined, in part by applying the previously described mass labeling method for the detection of ketones/aldehydes in SRFA following different irradiation protocols. SRFA was irradiated employing two long-pass filters (418 nm and 320 nm) and then reduced with $NaBD_4$.

Irradiation resulted in a loss of high molecular weight species and a loss in absorption that was linked to optical parameters, $E2/E3$ ratio and spectral slope,

to reveal additional information regarding photodegradation of CDOM. To examine if irradiated terrestrial material appears similar to marine material and reveal possible explanations for the differences seen in source material from riverine to marine DOM, irradiated samples were compared to two open ocean samples. Reduction was also employed on all samples to observe how photodegradation can effect ketone/aldehyde- containing species that are largely responsible for long wavelength absorption and gain insights into the composition and contribution to the optical properties of the bulk (irradiated) DOM.

5.2 Materials and Methods

5.2.1 Reagents and Materials

Suwannee River fulvic acid (SRFA) was purchased from International Humic Substances Society, (Standard II, catalog number: 2S101H). Ammonium hydroxide (NH_4OH), perchloric acid ($HClO_4$), sodium borodeuteride ($NaBD_4$; 98% deuterium, 90% purity), and quinine sulfate were purchased from Sigma Aldrich. Purified water was acquired from a Milli-Q purification system (Millipore). High purity methanol (MeOH) was purchased from Honeywell Burdick and Jackson. Bond Elut PPL cartridges, a proprietary functionalized styrene-divinylbenzene polymer, (100 mg, part number 12105004) were purchased from Agilent. Turmeric 110 Analytical Test Strips were purchased from Scientific Equipment of Houston (Lot 4012). For irradiation studies long-pass filters (418 and 320 nm) were employed.

5.2.2 Sample Preparation of SRFA Solution for Irradiation and Reduction

Solutions of SRFA were prepared in Milli-Q water (10 mg/L) and filtered through a preconditioned 0.2 μm nylon filter. Optical measurements were taken at pH$\sim$7 before samples were split into aliquots for untreated controls and irradiation and reduction experiments (scheme of sample preparation is depicted in Appendix F Figure F.1).

Irradiation For each aerobic polychromatic irradiation, IRR418 was irradiated for 54 hours employing a 418 nm long pass filter and IRR320 was irradiated for 33 hours employing a 320 nm long pass filter, duplicates of 1-cm quartz cuvettes were place side by side and irradiated with a 300 W xenon lamp (filter absorption spectra are provided in Appendix F Figure F.2). To remove infrared radiation, a 20 cm water jacket was placed between the lamp and the cuvette. Absorbance measurements were taken throughout the irradiations to monitor the change of absorbance over time (Appendix F Figure F.3). To be comparable to the reduction irradiated samples were also extracted by PPL cartridges.

Borodeuteride Reduction Borodeuteride reduction was carried out as described previously Chapter 4 section 4.2.2. Briefly, aliquots of untreated and previously irradiated samples were treated with $\sim$50- fold mass excess of $NaBD_4$ relative to SRFA mass. Untreated (UNT) and $NaBD_4$ reduced samples (R50x, as in Chapter 4), RED IRR418, and RED IRR320) were kept open to air, at room temperature, and in the dark for 24 hours to allow for reduction. After 24 hours, optical mea-

surements were taken at pH 10 and again at pH 7 after adjusting the pH with 1N $HClO_4$. Samples were then adjusted to pH 2 with 1N HClO4 for extraction and to remove any residual sodium borodeuteride. Solid phase extraction (SPE) was used to purify samples by removing excess borate salts [81, 96]. Half of the DOM methanol eluate was stored for FT-ICR MS analysis at -18°C. The other half of the eluate was dried using a gentle N_2 stream and then dissolved in purified water for optical measurements (pH∼7 and pH∼10). Turmeric test strips were used to confirm the absence of residual borate.

5.2.3 Optical Measurements for SRFA: Irradiation and Reduction

Apparatus for Optical Measurements. For instruments employed for absorbance and fluorescence measurements see Appendix C.1 section D.1. Setup for irradiation studies Appendix F Figure F.4.

Comparing pre- and post- extraction material for irradiation study. As previously described in Chapter 4 optical measurements were taken on samples prior to and following (Post) SPE while mass spectra were collected exclusively on the Post-SPE samples. As expected, only minor differences were observed since SRFA is already an extract from a similar type of support (XAD-8 resin). In Appendix F Figure F.5 the absorbance spectra of pre vs post samples (normalized at 300 nm) are shown.

5.2.4 Mass spectrometric data acquisition, pre-processing, molecular formula assignments, and identification of reducible species in SRFA and irradiated SRFA

Mass spectrometric data acquisition and pre-processing for SRFA and irradiated SRFA. Mass spectrometric data acquisition and pre-processing details for untreated (UNT), borodeuteride-reduced, and irradiated *and* borodeuteride reduced (50-fold mass/mass SRFA) samples are in Chapter 3 section 3.2. For this study the mass range of 200-700 m/z was analyzed.

Molecular formula assignments and identification of reducible species in SRFA and irradiated SRFA. Data processing and analyses for molecular formula assignments and identifying ketone/aldehyde reducible compounds were undertaken using custom-written MATLAB functions described in detail in Chapter 3 section 3.3 (for MATLAB code see Appendix H). As in Chapter 4, including additional heteroatoms in assignments (formulae containing up to 2 N or 1 S) did not significantly increase the amount of the total peak intensity (TPI) able to be assigned. Further, including more than 2 D atoms did not show significant increases in reduction for reduced samples but rather increased false positives in untreated samples (Appendix F Figure F.6). Therefore, only CHO formulae and up to 2 D atoms in the 200-700 m/z range were analyzed for the subsequent analysis of untreated, irradiated, reduced, and combined irradiated *and* reduced samples.

5.3 Results and Discussion

5.3.1 Effect of Irradiation on the Optical Properties

The absorption changes due to irradiation were investigated. As described previously, the unirradiated SRFA CDOM absorption spectra decreased exponentially with increasing wavelength and was fairly featureless. Irradiation of SRFA with the 320 nm filter resulted in a loss of absorption, similar to that reported for a 59 hr irradiated sample in Sharpless and Blough 2014 [40] and a 170 hour (305 cut-off filter) in Kujawinski et al. 2004 [116]. The irradiation of IRR320 started to reveal some shoulders (at ∼225 and 275 nm) to the usually unstructured SRFA CDOM spectrum (Figure 5.1). The greatest loss of absorption due to irradiation was observed in the visible region (∼40% for IRR418 and ∼60% IRR320 at 400 nm) (Figure 5.1 b). Irradiation also resulted in an increase in spectral slope (previously discussed in Chapter 1 and equation provided 1.1a), which suggests high molecular weight species are converted into lower molecular weight species due to photodegradation (Figure 5.1 d).

As well, the $E2/E3$ ratio (ratio of absorbance at 250 nm/365 nm) increased after irradiation (Figure 5.1 b), also suggesting irradiation decreases the aromaticity and molecular size of the DOM sample (as $E2/E3$ increases, aromaticity and molecular size decrease [47]). The molecular weight dependence of the $E2/E3$ ratio is thought to be related to the increased probability of electronic interactions (chromophore-chromophore interactions) of large DOM molecules responsible

for the long wavelength absorption as described by the electronic interaction (EI) model (discussed in Chapter 1); therefore, the increase of $E2/E3$ suggests that the chromophore-chromophore interactions (between electron donating and electron accepting moieties) are being destroyed by irradiation as a result of a loss of high molecular weight species and photooxidation of donor/acceptor moieties in DOM [117].

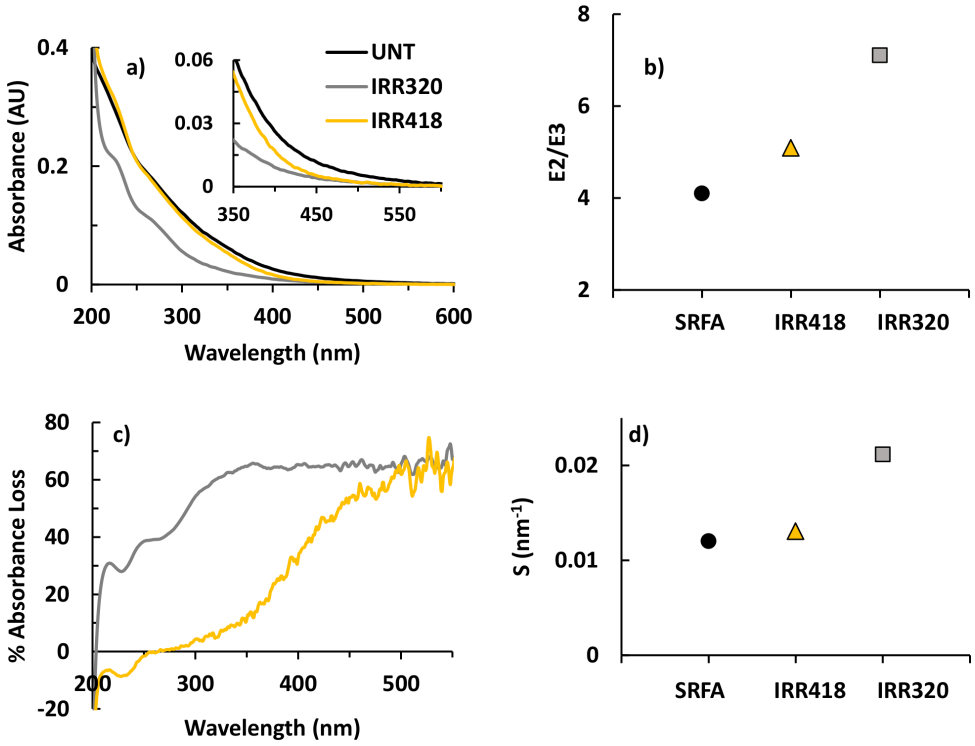


Figure 5.1: Absorption spectra at pH7 (prior to solid phase extraction with PPL). **a)** Absorption of untreated and irradiated samples (1cm cuvette). Inset: Long wavelength absorption -expanded region from 350 to 600 nm. **b)** $E2/E3$ ratios. **c)** Percent loss of absorption after irradiation. **d)** Spectral slope from 275-295 nm ($S_{275-295}$). Untreated-UNT (black) Irradiated samples-IRR418 (yellow) and IRR320 (gray). IRR418 was irradiated for 54 hours employing a 418 nm long pass filter and IRR320 was irradiated for 33 hours employing a 320 nm long pass filter.

5.3.2 Effect of Irradiation on the Mass Spectrometric Features

The molecular composition data obtained by ESI FT-ICR MS revealed a decrease in average molecular weight as a result of irradiation, which has also been seen in other MS photodegradation studies [12, 105, 116] as well as a size-exclusion chromatography study [117]. IRR418 average molecular weight decreased ~20 m/z and IRR320 average molecular weight decreased ~60 m/z compared to the untreated SRFA sample (Table 5.1). The decrease in average molecular weight is due to a loss of higher molecular weight species (>500 m/z) as well as an increase of lower weight species (<~400 m/z, especially seen in IRR320) (Figure 5.2 and Appendix F Figure F.7). The decrease in molecular weight and degradation into smaller weight compounds could be attributed chemical alterations by reactive oxygen species and loss of CO_2, CO, and H_2O [105].

Table 5.1: Mass spectral peak distributions (200-700 m/z)

Sample	$\mathbf{M_N}^{a,b}$	$\mathbf{M_W}^{a,c}$	$\mathbf{M_N/M_W}$	$\mathbf{TPI}^{a,d}$	$\mathbf{DBE}^{e}$	$\mathbf{H/C}^{e}$	$\mathbf{O/C}^{e}$
SRFA	463	487	0.95	1.4E+11	12	1.12	0.57
IRR418	445	464	0.96	3.4E+10	11	1.16	0.54
IRR320	406	425	0.96	2.5E+10	10	1.22	0.53

[a]singly-charged, blank subtracted, error < 0.4 ppm; [b] M_N: number-averaged m/z; [c] M_W: weight-averaged m/z; [d] TPI: total peak intensity; [e] intensity-weighted

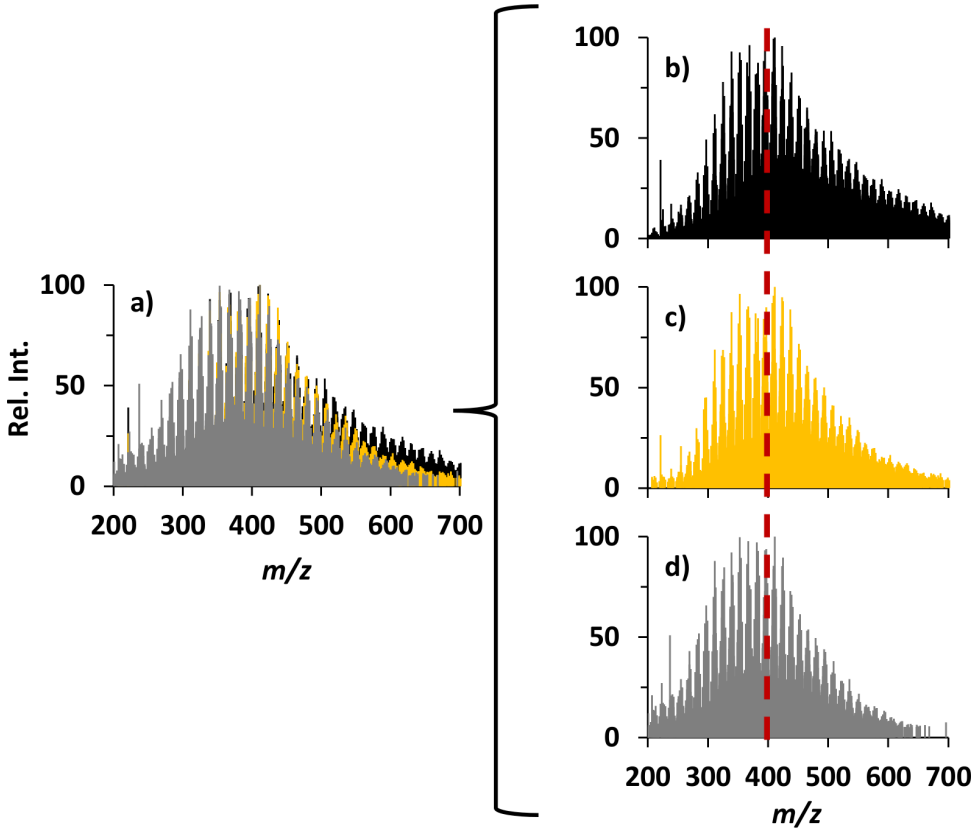


Figure 5.2: Negative ion ESI FT-ICR mass spectra (200-700 m/z) of peaks assigned CHO molecular formulae. a) Untreated sample (black), overlaid with irradiated samples: IRR418 (yellow) and IRR320 (grey); b) Untreated sample; c) IRR420-irradiated for 54 hours with 418 nm long pass filter; d) IRR320-irradiated for 33 hours with 320 nm long pass filter.

Overall, after irradiation the average DBE decreased, average H/C ratio increased, and average O/C ratio decreased (Table 5.1). In untreated SRFA there are DBE >20, in the >∼500 m/z range, which are not present in the irradiated samples (Figure 5.3). Figure 5.4 shows the decrease in relative abundance of DBE after irradiation, specifically for higher DBE, but also the increase in relative abundance of lower DBE after irradiation (as well as shift from high DBE-O to low DBE-O

values which was also observed by Gonisor et al. 2009 [109] (Appendix F Figure F.8)); this complements the mass distribution data as well as aligns well with the increase of spectral slope and increase of $E2/E3$ after irradiation suggesting a loss of unsaturated/aromatic compounds and loss of higher molecular weight compounds. The relative abundance of DBE was calculated as

$$DBE_{Rel.Abund.} = \frac{\sum DBE_{intensity}}{\sum intensity} \tag{5.1a}$$

where $\text{DBE}_{intensity}$ is the peak intensity of a specific double bond equivalent and intensity is the peak intensity.

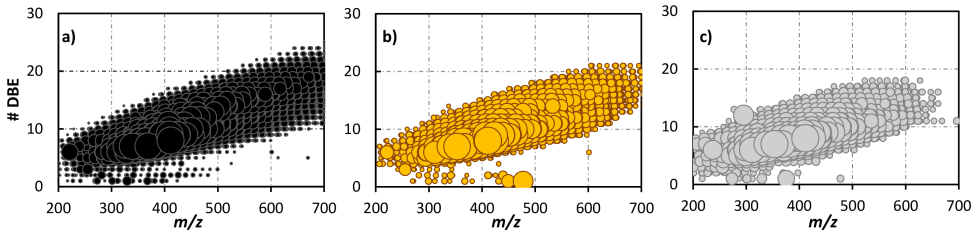


Figure 5.3: DBE vs m/z of untreated and irradiated SRFA samples; a) Untreated (black); b) IRR418 (yellow); c) IRR320 (gray).

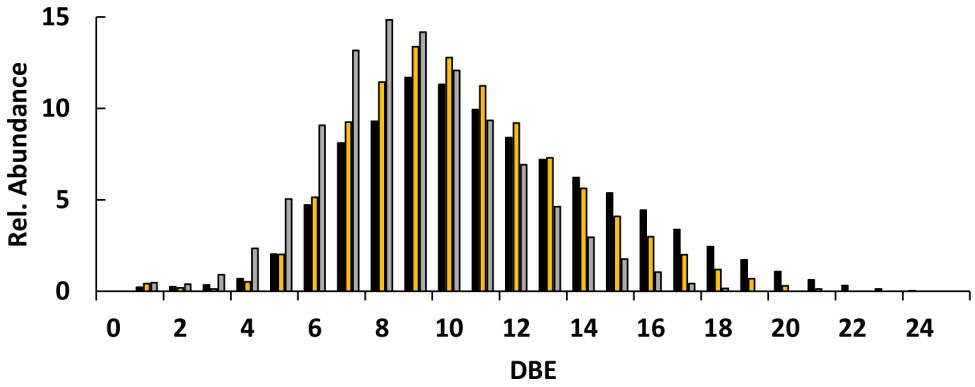


Figure 5.4: Relative abundance of DBE. Untreated sample black, irradiated samples: IRR418 (yellow) and IRR320 (gray).

Van Krevelen plots were created to visualize the composition and the increase/decrease of peaks after irradiation (Figure 5.5). A slight increase in high H/C ratios can be observed for IRR320 in the 200-400 m/z range (indicating an increase in aliphatic content), while the species that were lost and/or decreased due to irradiation were at lower H/C ratios (H/C$\sim$<0.75) (indicating a loss of unsaturated hydrocarbons) and higher O/C ratios for the 400-600 m/z range, and were at both high and low H/C and O/C for the 600-700 m/z range.

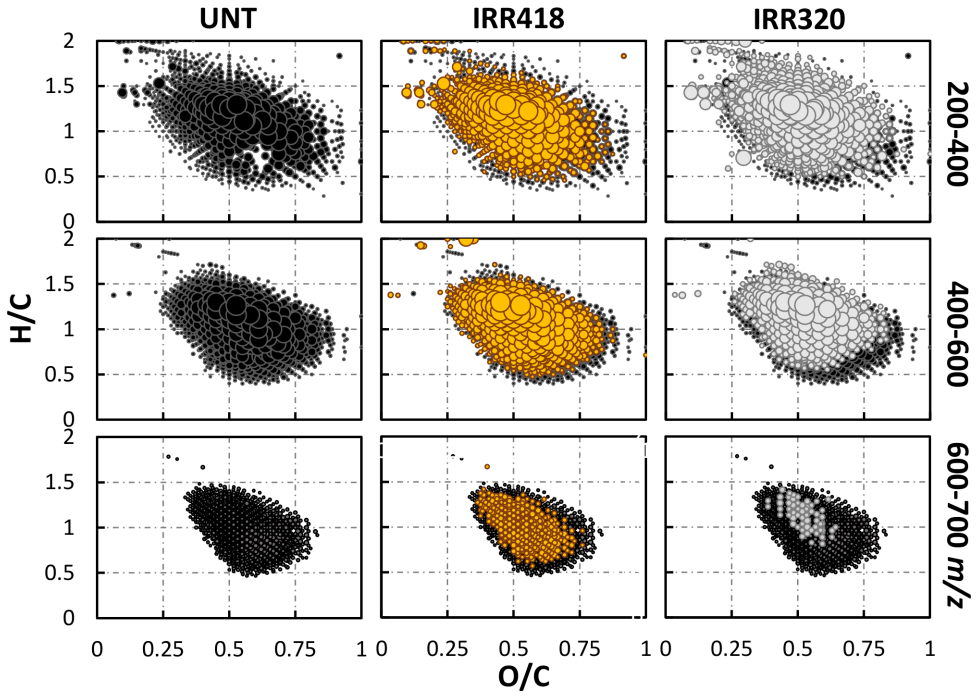


Figure 5.5: Van Krevelen plots of CHO formulae identified in untreated and irradiated SRFA (UNT, IRR418, and IRR320) in three mass ranges (200-400, 400-600, and 600-700 m/z). Untreated sample shown in black and is also overlaid with IRR418 (yellow) and IRR320 (grey) respectively. Size of the circle is relative to the peak height.

A few previous studies have proposed ideas for the possible losses of peaks after irradiation. As noticed by Stenson et al. 2003 [12] and Kujawinski et al. 2004 [116], when looking at some expanded nominal mass regions, every other peak is missing in the irradiated samples; an example for this study is shown at 442 m/z (Figure 5.6). These missing peaks differ by multiples of 0.036 Da (difference between CH_4 and O); Kujawinski et al. 2004 suggested this mass difference may occur from the replacement of a methyl group by a ketone or aldehyde [116].

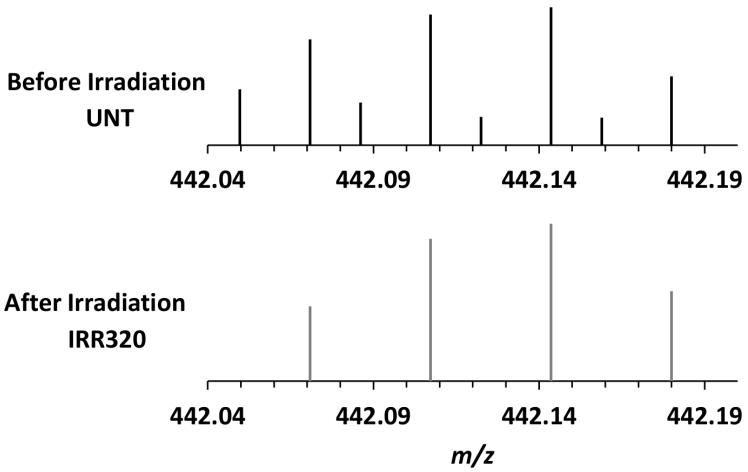


Figure 5.6: Effect of irradiation on SRFA at nominal mass 442 *m/z*. Negative ion ESI FT-ICR mass spectra for Top) Before irradiation, untreated (UNT) and Bottom) after irradiation (IRR320).

The study by Kujawinski et al. 2004, observed that the missing peaks had high DBE and lower oxygen values, while the peaks that remained (are not affected by irradiation) had low DBE and high oxygen values [116]); these findings were not fully consistent with the observations in this study. This study found peaks with both high DBE and lower oxygen values *and* low DBE and higher oxygen values were lost upon irradiation (Figure 5.8).

Kujawinski et al., hypothesized that 1) high DBE species which are conjugated are getting destroyed by the longer wavelength polychromatic light and 2) the compounds with low DBE and high oxygen are less photoreactive and therefore remain after irradiation [116]. The second statement does not align with the observations of this study since low DBE and high oxygen formulas were observed to be also lost in this study. As well, the first statement only partly aligns with the observations of this study because although unsaturated/aromatic compounds were

observed to be lost after irradiation with longer wavelength (visible light), IRR418, there was smaller loss of high DBE and lower oxygen formulas observed for IRR418 than IRR320 (for the peaks lost after irradiation with the 418 nm filter, DBE # > Oxygen # = 42% and DBE # < Oxygen # = 58% and for the 320 nm filter, DBE # > Oxygen # = 54% and DBE # < Oxygen # = 46%). To note, the irradiated sample in Kujawinski et al. 2004 was irradiated for 170 hours with a 305 nm filter and in this study IRR320 was irradiated for 33 hours; therefore, some of the differences noticed between the two studies could be due to differences in photon doses/irradiation time, as the SRFA sample in Kujawinski et al. was irradiated with UVB radiation (300 to 320 nm), UVA radiation (320 to 400 nm), and visible radiation (400 to 700 nm) and lost ∼80% of absorption after irradiation. In this study IRR320 was irradiated with only UVA and visible radiation and lost ∼60% respectively (future studies could monitor the changes in absorption and mass spectra with time while using the same filter to better understand the changes in molecular composition with both increasing exposure in terms of time and radiation).

Another irradiation study, by Gonsior et al. 2009 [109], observed the loss of H/C and O/C ratios from a river estuary sample that supports the idea of high DBE and low oxygen formulas transforming into more saturated formulas after irradiation and noticed preferential loss of highly aromatic formulas after a 21 hr irradiation with simulated sunlight. Better understanding photobleaching/photodegradation can help for understanding the distribution of DOM in the aquatic environment (from terrestrial to marine samples). Studies have suggested that some of the material that makes up marine DOM is highly degraded lignin material [106]. To gain

insights into if the photodegradation of terrestrial material appears similar to marine material, this study compares terrestrial sample SRFA IRR320 to a surface and deep ocean sample from the North Pacific Ocean (NPO) (sample preparation and further discussion of NPO samples is in Chapter 6). Figure 5.7 shows both NPO samples have very few peaks below an H/C of 1 and are centered around H/C $\sim$ 1.25 and O/C $\sim$ 0.5 for all mass ranges (200-700 m/z), which is comparable to the higher H/C ratio (1.22) observed for the IRR320 vs the untreated SRFA.

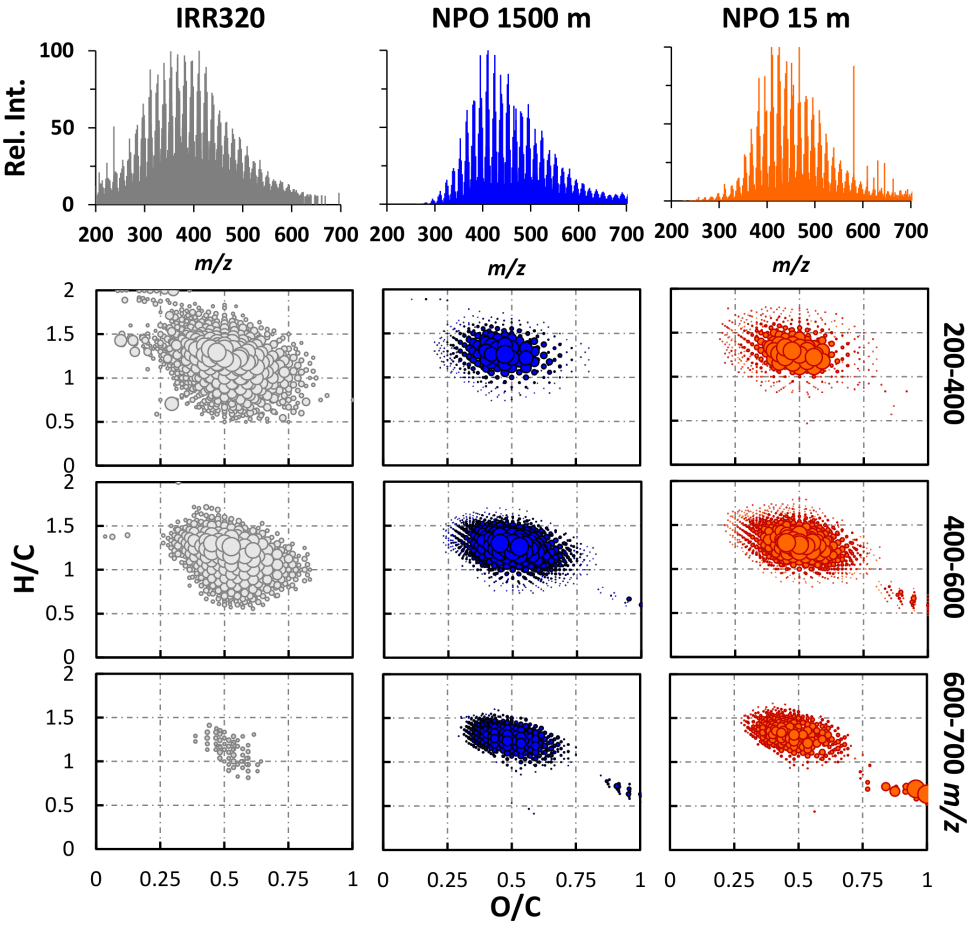


Figure 5.7: Top) Negative ion ESI FT-ICR mass spectra from 200-700 *m/z*. : SRFA IRR320 (grey), North Pacific Ocean (NPO) depth of 1500 m (blue) and 15 m-surface sample (orange). **Bottom)** Van Krevelen plots of CHO formulae identified in three mass ranges (200-400, 400-600, and 600-700 *m/z*) for SRFA IRR320 (grey), NPO 1500 m (blue) and NPO 15 m (orange). The high O/C peaks in the ocean samples are artifacts. Size of the circle is relative to the peak height.

To gain better insight into the loss of peaks possibly due to irradiation/photobleaching, Van Krevelen diagrams were created for peaks that were in the untreated sample but *not* in the respective irradiated samples as well as peaks that were in the NPO deep ocean sample but *not* in the surface sample (the 15 m sample should be more photobleached than the 1500 m sample due to 1500 meters being well below

the surface mixed layer, which reaches down to $\sim$ 10 m in the summer and 300-400 m in the winter [118]) (Figure 5.8). Loss of low H/C peaks can be clearly seen for both the irradiated and marine surface sample. IRR320 has the greatest loss at masses >400 m/z as does the surface sample. This result suggests that terrestrial material can possibly transform into more marine-like material after photobleaching and degradation processes through alteration of the chemical composition and loss of unsaturated/aromatic compounds.

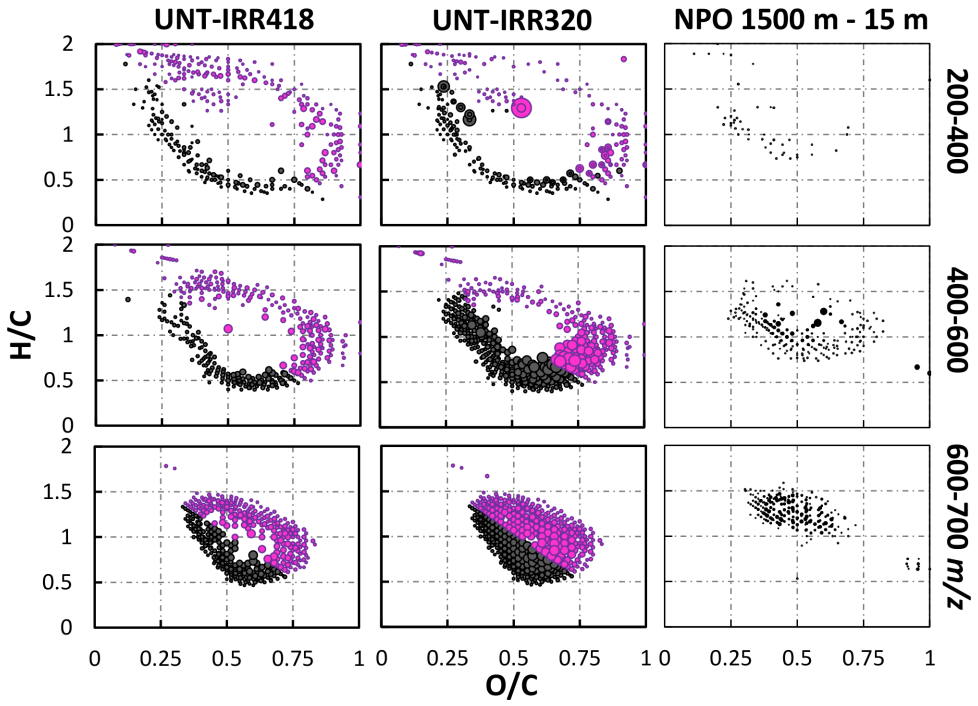


Figure 5.8: Van Krevelen plots comparing what is lost upon irradiation of SRFA and from surface to deep ocean water in three mass ranges (200-400, 400-600, and 600-700 m/z). For UNT-IRR418 and UNT-IRR320: shown are peaks in the untreated sample that are not in the respective irradiated sample (peaks lost due to irradiation). Higher DBE, lower oxygen (DBE # > Oxygen #) in black; Low DBE, higher oxygen (DBE # > Oxygen #) in pink. For NPO 1500 m - NPO 15 m: shown are peaks that are in the 1500 m sample (deep water) but not in the 15 m sample (surface water).

5.3.3 Effect of Reduction and Combined Irradiation *and* Reduction of SRFA on the Optical Properties

The absorption changes after reduction for both untreated and irradiated SRFA were investigated. For all samples, reduction by $NaBD_4$ resulted in loss of absorption at wavelengths > 250 nm and the greatest loss of absorption was in the visible region (Figure 5.9, Appendix F Table F.1). R50x and RED IRR418 have a very similar loss in absorption across all wavelengths; however, there is a slight increase in percent loss of absorption in the visible region for RED IRR418 compared to R50x (Figure 5.9 g&h), while in contrast RED IRR320 has a similar loss up to $\sim$325 nm and decreased there after (compared to both R50x and RED IRR418). IRR320 had the highest overall percent loss in absorption, compared to IRR418, when considering the loss due to both irradiation *and* reduction, $\sim$ 65% vs $\sim$ 40% in the 250-500 nm range (Appendix F Table F.1). Most of the absorption loss is due to irradiation for IRR320, while for IRR418, the absorption loss is from reduction (Figure 5.9). IRR320, having a smaller percent loss in absorption due to reduction, than both R50x and RED IRR418, suggests that the irradiation with the 320 nm filter destroyed ketone/aldehyde acceptors. On the other hand, IRR418 increased after reduction in the visible region which could possibly be due to photo-oxidation and increase in aldehyde/ketone species (replacement a methyl group with a ketone or aldehyde as suggested by Kujawinski et al. 2004 [116]), that are then able to be reduced by $NaBD_4$ as well irradiation at this wavelength could possibly break weak interactions between complexes that then make ketone/aldehyde moieties more

easily accessible (for reduction).

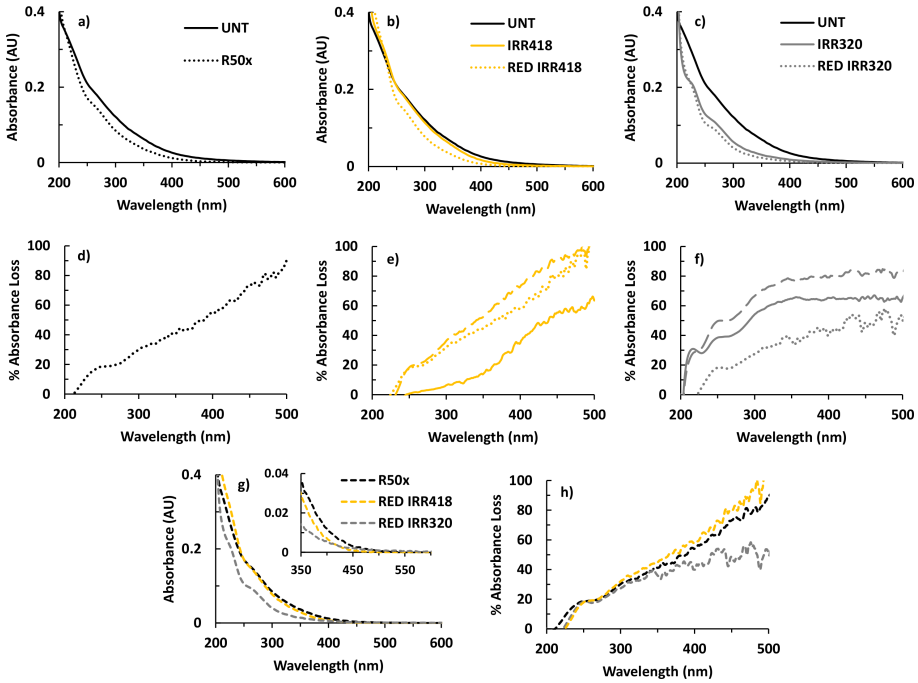


Figure 5.9: a,b,c) Absorption and d,e,f) percent of absorption loss for reduced, irradiated, and combined irradiated *and* reduced samples compared to untreated SRFA (control sample). Untreated, UNT (solid black line); R50x (dotted black line), irradiated samples, IRR418 (solid yellow line), IRR320 (solid gray line); reduced samples, RED IRR418, (dotted yellow line), RED IRR320, (dotted gray line); percent absorption loss from both irradiation + reduction (dashed line); g&h) comparison of all reduced samples.

5.3.4 Effect of Reduction and Combined Irradiation *and* Reduction of SRFA on the Mass spectrometric Features

The reduction of SRFA was described in detail in Chapter 4 and the mass labeling method used to identify ketone/aldehyde-containing species was described in Chapter 3; here reduction (50-fold mass $NaBD_4$/mass SRFA) is applied to irradi-

ated SRFA samples. UNT, IRR418 and IRR320 were used as controls by searching for the unique $M+3n$ (n_{max}=2) in their respective (not-reduced) peak lists. The controls resulted in falsely identifying less than 1% of CHO molecular formulae as reduced species (Appendix F Figure F.6, [0,0,2]), which resulted in very low $\%I_{total}$ (sum of $\%I_{1O}$, $\%I_{1B}$, $\%I_{2O}$, and $\%I_{2B}$), indicating that this method generates very few false positives for both the untreated and irradiated samples.

As also seen in Chapter 4, there was a slight decrease in the average m/z after reduction of the samples (for explanation see Chapter 4 section 4.3.1). The average DBE decreases and H/C ratio increases after reduction for all samples (compared to their untreated counterpart), while the O/C ratio does not significantly change (Table 5.2).

Table 5.2: Mass spectral peak distributions after reduction (200-700 m/z)

Sample	$\mathbf{M_N}^{a,b}$	$\mathbf{M_W}^{a,c}$	$\mathbf{M_N/M_W}$	$\mathbf{TPI}^{a,d}$	$\mathbf{DBE}^{e}$	$\mathbf{H/C}^{e}$	$\mathbf{O/C}^{e}$
R50x	448	466	0.96	1.5E+11	11	1.23	0.53
RED IRR418	433	452	0.96	1.2E+11	10	1.28	0.50
RED IRR320	400	419	0.96	1.1E+11	9	1.35	0.52

[a]singly-charged, blank subtracted, error < 0.4 ppm; [b] M_N: number-averaged m/z; [c] M_W: weight-averaged m/z; [d] TPI: total peak intensity; [e] intensity-weighted

The distribution of the reduced peaks vs m/z is shown in Figure 5.10. The highest amount of total reduction for all samples occurs between 350-500 m/z. There is a slight overall decrease in $\%I_{total}$ after irradiation and there is a substantial loss of 1O and 1B peaks in the irradiated samples at masses >∼450 m/z (Figure 5.10). For IRR320, there is an increase in reduction in the 200-400 m/z range for $\%I_{total}$ (mostly from $\%I_{1B}$ and $\%I_{2B}$), compared to both R50x and RED IRR418 (Figure

5.11), which is most likely due to the shifted mass distribution of IRR320 to lower m/z values.

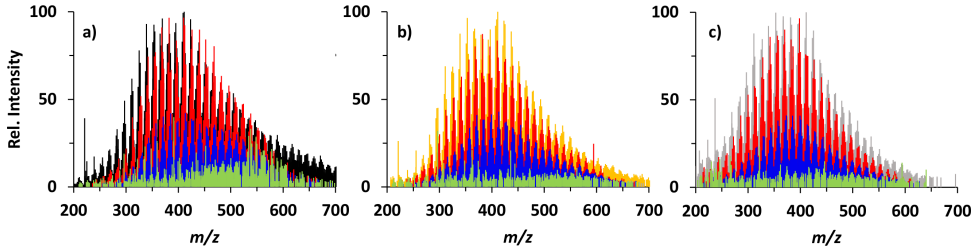


Figure 5.10: Negative ion ESI FT-ICR mass spectra (200-700 m/z) of peaks assigned CHO molecular formulae for reduced samples (50-fold mass $NaBD_4$/mass SRFA). **a)** Untreated sample (black), overlaid with reduced sample (R50x); **b)** Irradiated 418nm long pass filter sample (yellow), overlaid with reduced sample (RED IRR 418); **c)** Irradiated 320nm long pass filter sample (grey), overlaid with reduced sample (RED IRR 320). Intensity of reduced peaks depicted on a, b, and c as 1O (green, M+3 only), 1B (red, both M+3 and M+6, showing intensity of M+3 peak), and 2B (blue, both M+3 and M+6, showing intensity of M+6 peak). 2O not shown (less than 1% of the peak intensity for all samples).

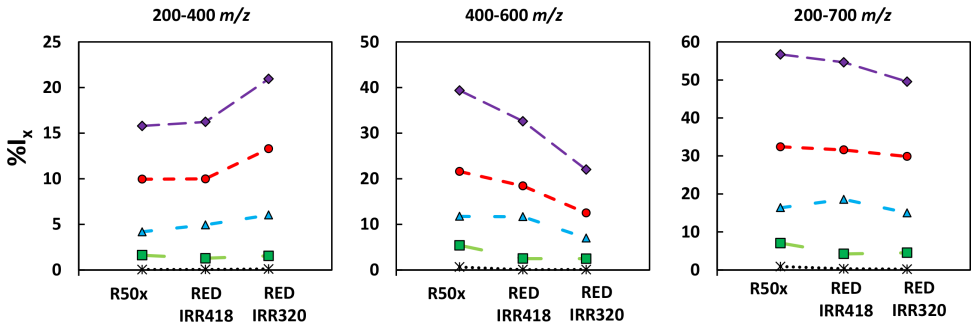


Figure 5.11: $\%I_X$ extent of reduction vs R50x and irradiated samples (IRR418 and IRR320). $\%I_X$ (intensity-weighted percent of ions from 200-400, 400-600, and 200-700 m/z). Total percent of reduced intensity (purple, $\%I_{total}$) allocated into four groups: $\%I_{1O}$ (green) % of intensity of M+3 only peaks), $\%I_{2O}$ (black) % of intensity of M+6 only peaks; $\%I_{1B}$ (red) both M+3 and M+6 found, showing % of intensity of M+3 peaks, $\%I_{2B}$ (blue) both M+3 and M+6 found, showing % of intensity of M+6 peaks.

Van Krevelen plots were constructed to visualize peaks arising from reduction of the untreated and irradiated samples. As also seen in Table 5.2, the H/C and O/C of the majority of all formulae were centered around H/C of slightly over 1 and O/C $\sim$ 0.5. Reduced species of all of the samples have a similar distribution across the Van Krevelen space (Figure 5.12) and a very small amount of reduction occurs on species >600 m/z for all samples as also observed in DBE vs m/z in Figure 5.13. Furthermore, the distribution, across the mass range of 200-600 m/z for 1B and 2B peaks are very similar for all reduced samples. 1B species have the highest intensity (larges circles) and are located near the center of the plots, showing that even after irradiation the reduction is still acting on highly related molecular formulas.

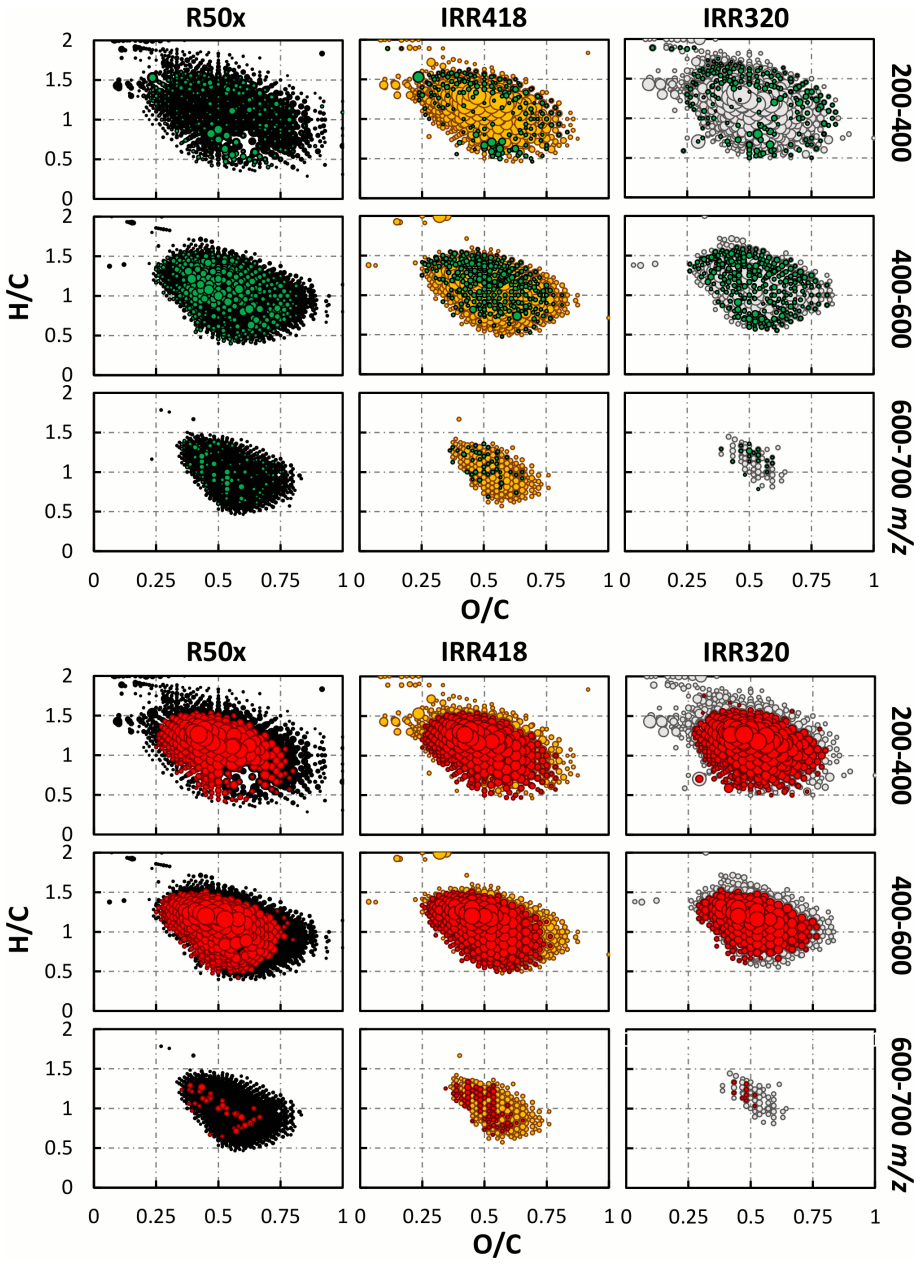


Figure 5.12: Van Krevelen plots of CHO formulae identified in untreated, irradiated, reduced, and irradiated and reduced SRFA (UNT, IRR418, IRR320, R50x, RED IRR418, and RED IRR320, respectively) in three mass ranges (200-400, 400-600, and 600-700 m/z). UNT (black), IRR418 (yellow) and IRR320 (grey), overlaid with their peaks arising from reduction. **Top)** 1O (green, M+3 only) and **Bottom)** 1B (red, both M+3 and M+6, showing intensity of M+3 peak). 2B both M+3 and M+6, showing intensity of M+6 peak shown in Appendix F Figure F.9. 2O not shown (less than 1% of the peak intensity for all samples). Size of the circle is relative to the peak height.

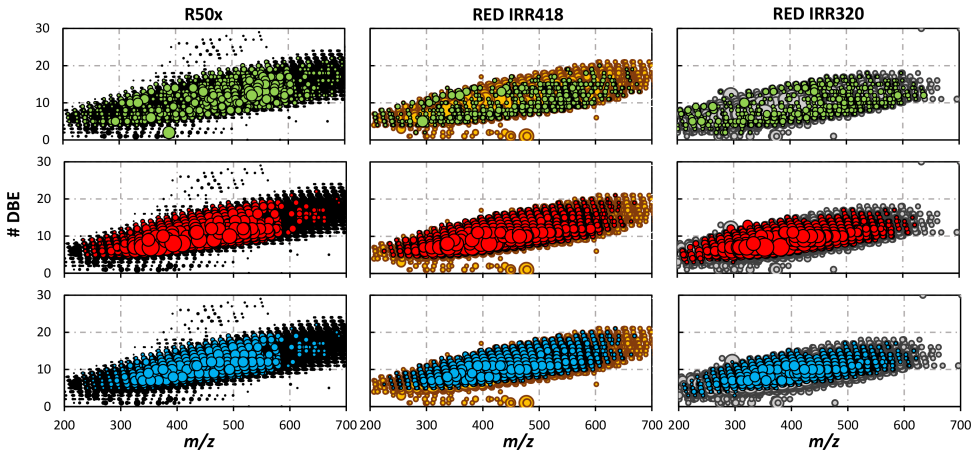


Figure 5.13: Double Bond Equivalents (DBE) vs m/z of CHO formulae identified in untreated, irradiated, reduced, and irradiated and reduced SRFA (UNT, IRR418, IRR320, R50x, RED IRR418, and RED IRR320, respectively) in three mass ranges (200-400, 400-600, and 600-700 m/z). UNT (black), IRR418 (yellow) and IRR320 (grey), overlaid with their peaks arising from reduction. **Top)** 1O (green, M+3 only) and **Middle)** 1B (red, both M+3 and M+6, showing intensity of M+3 peak). **Bottom)** 2B both (blue, both M+3 and M+6, showing intensity of M+6 peak). 2O not shown (less than 1% of the peak intensity for all samples). Size of the circle is relative to the peak height.

5.3.5 Relation of Identified Ketone/aldehyde-containing Species in Reduced and Combined Irradiated *and* Reduced SRFA to the Observed Optical Properties After Reduction

$\%I_{total}$ was compared to the absorbance loss for R50x, RED IRR418, and IRR320 in Table 5.3. R50x and IRR418 had very similar $\%I_{total}$ and percent loss of absorption across all three wavelength ranges investigated (250-300, 300-400, and 250-500 nm), while RED IRR320 had a lower $\%I_{total}$ and smaller absorption loss. This resulted in a high correlation between $\%I_{total}$ and the percent loss of absorption

due to reduction within three wavelength regions shown in Figure 5.14 (though there are only three samples being compared and therefore to be able to draw more conclusions on how strong the correlation is between $\%I_{total}$ and absorbance upon reduction, more irradiated samples should be compared).

IRR320, had a higher loss of absorption due to irradiation than reduction while IRR418 was the opposite, with higher amounts of absorption loss due to reduction than irradiation. The 320 nm filter, has greater amounts of irradiation into the UV-A region while the 418 nm filter, subjects the sample to only the visible region (Appendix F Figure F.2). Therefore, the 320 nm filter allows for more higher energy UV light which could produce more direct and indirect photochemistry in the IRR320 sample that is not observed in the IRR418. The higher mass region, where most of the species lost due to irradiation for both IRR418 and IRR320 (IRR418-average weight lost 541 m/z, IR320-average weight lost 566 m/z) is not where a significant amount of reduction is observed for any of the samples. Though, as stated above, RED IRR320 showed a decrease in the amount of reduction observed by mass spectrometry as well as optically. Therefore, even though most of the formulas lost are at higher m/z values, losses were still observed across all masses. The potential photochemical reactions occurring due to UV-A light in IRR320 and destroying ketone/aldehyde acceptors could be responsible for the decrease in the amount of reducible species by $NaBD_4$ and therefore the overall lower reduction observed in RED IRR320.

Table 5.3: $\%I_{total}$ vs percent absorbance loss due reduction

sample	$\%I_{total}$	% Abs loss (from reduction)		
	200-700 m/z	250-300 nm	300-400 nm	250-500 nm
R50x	57	22	39	32
RED IRR418	55	22	42	32
RED IRR320	50	20	35	27

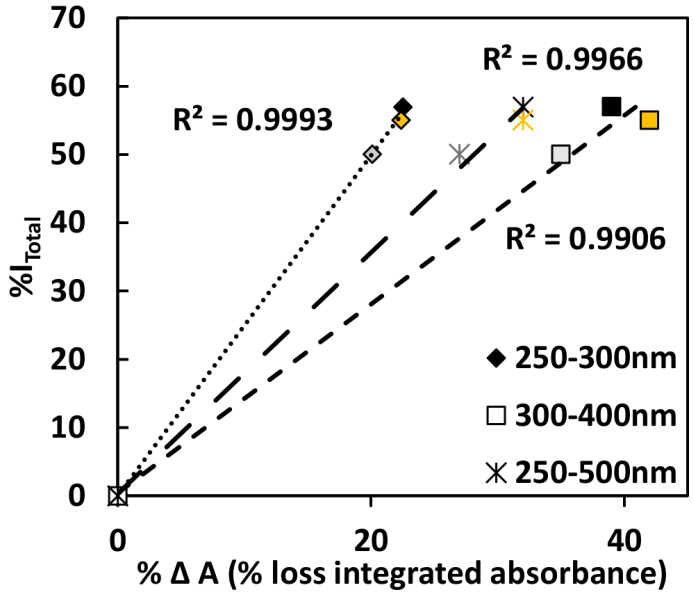


Figure 5.14: Percent of integrated absorbance loss upon reduction with $NaBD_4$ relative to untreated and respective irradiated samples (at pH 7) across three wavelength ranges correlated to $\%I_{total}$. 250-300 nm (diamonds), 300-400 nm (squares), and 400-500 nm (asterisks). R50x (black), IRR418 (yellow), IRR32) (grey).

5.4 Conclusions

This irradiation study correlated the observed changes in absorbance with the respective changes observed in the mass spectra, revealing insights into the structural components in SRFA that were lost or decreased due to irradiation. The Van

Krevelen plots showed a loss of lower H/C ratio peaks after the samples were irradiated which correlated to mostly unsaturated/aromatic species as seen also by a decrease in high DBE and increase in low DBE after irradiation. Furthermore, to assess if photodegradation makes terrestrial material appear more like marine samples, IRR320 was compared to open ocean samples. The open ocean samples also had a lack of low H/C ratios like IRR320. In addition, the borodeuteride mass labeling method was applied to the irradiated samples, which showed that irradiation can potentially increase or decrease the amount of reduction depending on the wavelength range of the irradiation. The 418 nm filter irradiated sample showed an increase in the percent of reduction which could be due to the photo-oxidation of ketones and aldehydes and which then were able to be reduced, while IR320, decreased in the percent of reduction (observed optically as well as by mass spectrometry), suggesting irradiation at higher energy (UV-A) can degrade ketone/aldehyde acceptors, removing them before reduction with $NaBD_4$.

Chapter 6: Comparison of the Contribution of Ketone- and Aldehyde-containing Species to the Composition and Optical Properties of DOM from Various Aquatic Locales

6.1 Introduction

Differences in the optical properties of DOM/CDOM from diverse aquatic environments have been observed in many studies [10, 31, 36, 50, 119], but how such optical properties depend on the source(s) and structure of DOM is still poorly understood [120]. Evidence has been provided for CDOM in rivers and estuaries to be derived primarily from lignin degradation products and possibly tannins [10–15]; however, in the open ocean away from the direct influence of rivers, the source(s) of CDOM remains controversial. Studies have provided evidence of a terrestrial input to the CDOM found in open oceans [10, 13, 16], while others have claimed that marine CDOM arises *in-situ* from marine biomass [17, 18].

This chapter employs the mass labeling method from Chapter 3 section 3.3 for the detection of borodeuteride-reducible ketone/aldehyde-containing species in DOM from different aquatic environments. The previous chapters, as well as prior work [16,103–105,107,108,115], have demonstrated the importance of ketone/aldehyde-

containing species to the optical properties and photochemistry of DOM samples. Therefore, applying this method to gain information on the optical as well as mass spectrometric properties of samples from different locations aims to help better understand the relationship between the source(s) and structure of DOM.

This focus of this Chapter is on relating the DOM/CDOM bulk optical properties and optical parameters ($E2/E3$ ratio and spectral slope) to the composition data obtained through mass spectrometry of eight samples from a variety of aquatic environments. This method is employed on reference material Suwannee River fulvic acid (SRFA) as well as CDOM extracts from the Delaware Bay, Delaware shelf break, Station ALOHA (North Pacific Ocean), and from the Equatorial Atlantic Ocean.

6.2 Material and Methods

6.2.1 Reagents and Materials.

Suwannee River fulvic acid (SRFA) was purchased from International Humic Substances Society (IHSS), (Standard II, catalog number: 2S101H). Ammonium hydroxide (NH_4OH), perchloric acid ($HClO_4$), sodium borodeuteride ($NaBD_4$; 98% deuterium, 90% purity), and quinine sulfate were purchased from Sigma Aldrich. Purified water was acquired from a Milli-Q purification system (Millipore). High purity methanol (MeOH) was purchased from Honeywell Burdick and Jackson. Bond Elut PPL cartridges, a proprietary functionalized styrene-divinylbenzene polymer, (100 mg, part number 12105004) were purchased from Agilent. Turmeric 110 Ana-

lytical Test Strips were purchased from Scientific Equipment of Houston (Lot 4012).

6.2.2 DOM Samples and Locations for Comparison

The DOM samples used in this study are from the Okefenokee Swamp (reference material SRFA), Delaware Upper and Mid Bay (DEL), the North Pacific Ocean (NPO), and the Equatorial Atlantic Ocean (EAO) (Map of locations Figure 6.1). SRFA is an XAD-8 resin extract while the rest of the samples are C_{18} extracts. As first established, deep open ocean samples are highly similar to one another; open ocean surface samples differ from the deep ocean samples, but are also highly similar to one another. Thus, representative surface and deep ocean samples were chosen for the comparative analysis: surface samples from both the NPO and EAO and deep samples from both NPO (1500 m) and EAO (St. 51 1000 m). The eight samples chosen for comparison and their respective descriptions are in Table 6.1 below (for a complete depth profile of samples from the NPO (11 depths) and four EAO samples see Appendix G section G.2).

Table 6.1: Sample locations and descriptions for comparison

Abbreviation	Location	Description
SRFA	Okefenokee Swamp	Reference Material
DEL U Bay	39° 37' 30" N, and 75° 34' 48" W	Upper Bay (2 m)
DEL M Bay	39° 18' 24" N, and 75° 22' 59" W	Mid Bay (2 m)
DEL Shelf	38° 00' N, and 75° 2' 49"W	Shelf Break (2 m)
NPO Surf	2° 45' N, and 158° 00' W	Surface (15) m
NPO Deep	2° 45' N, and 158° 00' W	AAIW (1486 m)
EAO Surf	5° 00' S, and 00° 00' E	Surface (5 m)
EAO Deep	5° 00' S, and 00° 00' E	Deep (1000 m)

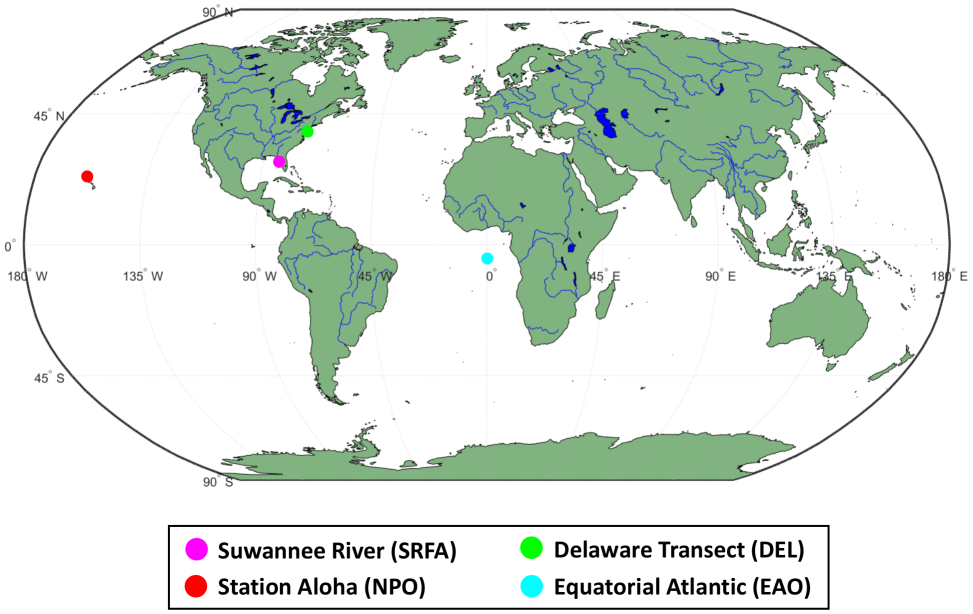


Figure 6.1: World map of sample locations (generated with MATLAB R2019b).

6.2.3 Sample Collection, Extraction, and Preparation

Sample collection and C_{18} extraction. For the collection of the natural water samples and extraction of DOM with C_{18} cartridges see Appendix G section G.1.

Sample preparation and reduction of DOM exacts. Aliquots of the samples were diluted to provide 10 mL solutions with absoption equivalent to a 10 mg/L solution of SRFA at 350 nm (in Millipore water). This solution was then split into two 5 mL aliquots (5 mL for the untreated control and 5mL for the reduction by $NaBD_4$) and were brought up to pH 10 with NH_4OH (1N). For reductions, extracts were reduced with ∼2.5mg $NaBD_4$ (which would equate to ∼50-fold mass excess for 10 mg/L SRFA solution). Untreated and borodeuteride reduced samples were stored in the dark and open to air at room temperature for 24 hours to allow for reduction. After 24 hours, optical measurements were taken at pH 10 and then again after adjusting the pH down to 7 with $HClO_4$ (1N). To remove the borate salts from the reduction, preconditioned PPL cartridges were employed. Untreated and reduced samples were adjusted to pH 2 with $HClO_4$ (1N) prior to loading them onto a PPL cartridge (conditioned with 5 mL MeOH and 5 mL water with 0.1% formic acid). The cartridges were then rinsed with 10 mL water (with 0.1% formic acid) and dried with a gentle nitrogen stream (15-20 min). The samples were eluted with methanol and the first 1 mL was collected. Half (0.5mL) of the DOM methanol eluate was stored for FT-ICR MS analysis, while the other half (0.5 mL) was dried using a gentle N_2 stream and dissolved in water for optical measurements (pH∼7 and

pH∼10). Turmeric test strips were used to test for residual borate. SRFA sample was prepared as described in Chapter 4 section 4.2 (R50x was used for comparison).

6.2.4 Mass Spectrometric Data Acquisition, Pre-processing, Molecular Formula Assignments, and Identification of Reducible Species in SRFA and DOM Extracts

Mass spectrometric data acquisition and pre-processing. Mass spectrometric data acquisition and pre-processing details for untreated and borodeuteride reduced SRFA and DOM extracts are provided in Chapter 3 section 3.2.

Molecular formula assignments and identification of reducible species in SRFA. Data processing and analyses for molecular formula assignments and identifying ketone/aldehyde reducible species were undertaken using custom-written MATLAB functions described in detail in Chapter 3 section 3.3 (for MATLAB code see Appendix H).

6.2.5 Optical Measurements for SRFA.

Apparatus for Optical Measurements. Instruments employed for absorbance measurements are described in Appendix C.1 section D.1.

Optical Measurements. Parameters and protocols used for absorbance measurements are detailed in Appendix C.1 section D.2.

6.2.6 Cluster Analysis for Similarity Comparison

Heatmaps were generated in MATLAB from the ESI FT-ICR MS data, using the Bray-Curtis dissimilarity measurement, which calculates the distance matrix of cluster analysis (Bray-Curtis distance matrix code from the Fathom toolbox by David Jones, University of Miami). This measurement has been shown to be reliable when applied to FT-ICR-MS data [119]. In this study, the Bray-Curtis dissimilarity calculation was applied using a presence/absence approach to the CHO molecular formula in both the untreated and reduced samples. The output from Bray-Curtis dissimilarity is between 0 and 1, with 0 meaning all samples share the same species and 1 meaning that they do not share any species. To observe how similar the samples are, the Bray-Curtis dissimilarity output values were subtracted from 1 and then multiplied by 100, to have the similarity as a percent, which is shown by the colorbar on the heatmaps in section 6.3.4.

6.3 Results and Discussion

6.3.1 Comparison of general mass spectrometric features.

The mass spectra of the eight samples compared in this study contained several thousand m/z ions from 150-800 m/z. However, only the 200-600 m/z range was analyzed because this range encompassed the majority of the total peak intensity (TPI) and because there is a higher chance for false positives due to the possibility of equivocal m/z assignments, as mentioned in Chapter 4, at m/z >600 (at a 1ppm

error) [91, 99]. The average molecular weights for all samples are between 400-500 m/z (M_N 456±11, M_W 468±9) (Table 6.2). The average molecular weights are highly based on ionization effiencey, EAO and NPO samples (both surface and deep) have very similar average molecular weights (M_N 463±2, M_W 473±2), while SRFA has the lowest average molecular weight. The ocean samples lack peaks <300 m/z) which shifts their mass distribution to be higher than SRFA, even though SRFA has a significantly higher amount of higher weight compounds than the ocean samples (Figure 6.2).

Table 6.2: Mass spectral peak distributions of untreated samples from different aquatic environments (200-600 m/z)

Sample	$\mathbf{M_N}^{a,b}$	$\mathbf{M_W}^{a,c}$	$\mathbf{M_N/M_W}$	$\mathbf{TPI}^{a,d}$	$\mathbf{DBE}^{e}$	$\mathbf{H/C}^{e}$	$\mathbf{O/C}^{e}$
SRFA	431	448	0.96	1.4E+11	10.4	1.07	0.54
DEL U Bay	457	472	0.97	2.6E+11	9.4	1.24	0.46
DEL M Bay	457	470	0.97	2.8E+11	9.5	1.24	0.46
DEL Shelf	452	465	0.97	2.8E+11	8.7	1.28	0.46
EAO Deep	463	474	0.98	3.5E+11	9.1	1.25	0.49
NPO Deep	463	473	0.98	2.0+E11	9.1	1.25	0.49
EAO Surf	460	471	0.98	3.8E+11	8.9	1.27	0.49
NPO Surf	465	475	0.98	4.0E+11	8.8	1.27	0.49

[a]singly-charged, blank subtracted, error < 0.4 ppm; [b] M_N: number-averaged m/z; [c] M_W: weight-averaged m/z; [d] TPI: total peak intensity; [e] intensity-weighted (CHO only)

The overall trend, seen from the eight samples, is the more riverine (and terrestrially influenced) the sample is, the higher the DBE and lower the H/C ratio while samples more off shore (into the open ocean) the lower the DBE and the higher the H/C ratio. These trends suggest that the marine samples contained more aliphatic species, while the riverine samples contained more aromatic species. Riverine DOM

has been shown to be derived from lignin and tannin degradation products which have a high aromatic content [10–15], while marine DOM has been shown to contain a significant contribution of carboxyl-rich alicyclic molecules (CRAM), which have higher H/C and lower O/C ratios [18, 70, 93]. As briefly discussed in Chapter 5 the difference between riverine and marine samples could in part be attributed to the photochemical degradation/photobleaching of terrestrial material during transit to offshore waters.

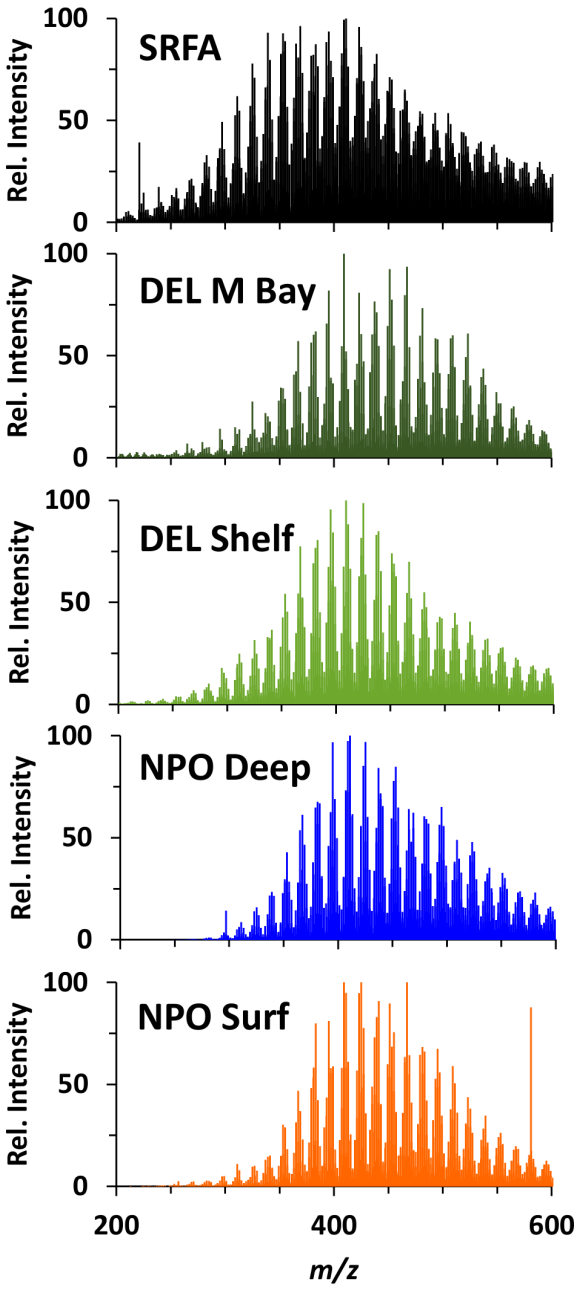


Figure 6.2: Negative ion ESI FT-ICR mass spectra (200-600 m/z) for comparison samples from different aquatic environments used for this study.

6.3.2 Comparison of Molecular Formula Assignments

In this study, for the C_{18} extracts, different allowances of ^{14}N, ^{32}S, and D were also tested (Appendix G Figure G.6). First, the amount of D-containing formulae was assessed. When including up to 2 D atoms, the % TPI of all of the untreated samples resulted in D-containing formulae that were <1% and all reduced samples had minimal equivocal assignments (<2% for D containing species); but when allowing for up to 3 D atoms, there was an increase of equivocal assignments (4-9%) with only minimal increases in %TPI (<4% for all reduced samples) and a slight increase in (false) identifications of D-containing species in the untreated samples. Therefore, like in Chapter 4, this result suggests it is unlikely that a significant amount of ionizable DOM species in the 200-600 *m/z* range have more than two $NaBD_4$ reducible species. When additional heteroatoms were included, a significant amount of equivocal formulae were found for $CHON_2S_1$ assignments in the C_{18} extracts. Analyzing the $CHON_1S_1$ assignments, there was less than a 9% increase for SRFA (as observed in Chapter 4), approximately a 15% increase was observed for both DEL U Bay and DEL M bay samples though the majority of the increase was from equivocal assignments, and ocean samples saw between a 9 and 15% increase which were mostly unequivocal assignments. Therefore, the majority of the TPI in the 200-600 *m/z* range for all samples consisted of CHO only species; however, the amount of CHO only species for the C_{18} extracts were found to be lower than in SRFA (as described above with C_{18} extracts having a higher increase in assignments with N and S inclusions than seen in SRFA). Though, as seen in Chapter 4

SRFA species containing only CHO species made up almost all (99%) of the TPI for the assigned reduced molecular formulae and in this study it was found that CHO only species for the C_{18} extracts accounted for ~98% of the reduced species. Also calculating $\%I_{total}$ with (up to 1 N and 1 S) and without heteroatoms, resulted in extremely similar values in reduction for each sample respectively (Appendix G Figure G.7). Therefore, compounds containing N and/or S contributed insignificantly to the reduced TPI and thus, only CHO only species and assignments with D up to 2 in the 200 – 600 m/z range were considered for subsequent analysis.

6.3.3 Comparison of Untreated and Reduced species

Untreated samples. As previously observed [10, 31, 50, 121], the CDOM absorption spectra, independent of sample location, decreased exponentially with increasing wavelength with no distinct peaks or shoulders with the exception of the deep ocean samples, which have a small shoulder at ~425 nm (this small band is also seen in NPO and EAO samples. NPO and EAO absorption spectra provided in Appendix G Figure G.2). Samples for this comparison study were normalized at 250 nm to visualize the change in spectral shape. Moving from riverine to marine samples, there was an increase in spectral slope ($S_{275-295}$) as well as an increase in the $E2/E3$ ratio (A_{250}/A_{365}) (Figure 6.4). As the $E2/E3$ ratio increases, aromaticity and molecular size have been shown to decrease [47]; therefore, the increase observed in the $E2/E3$ for marine samples, suggest the marine samples are less aromatic than the riverine samples, which is consistent with NMR data [47, 122].

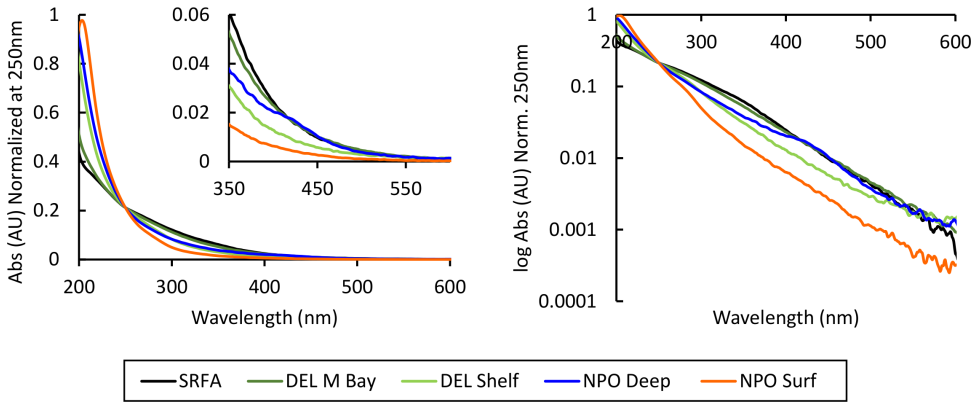


Figure 6.3: Absorption spectra of five of the untreated comparison samples normalized at 250 nm.

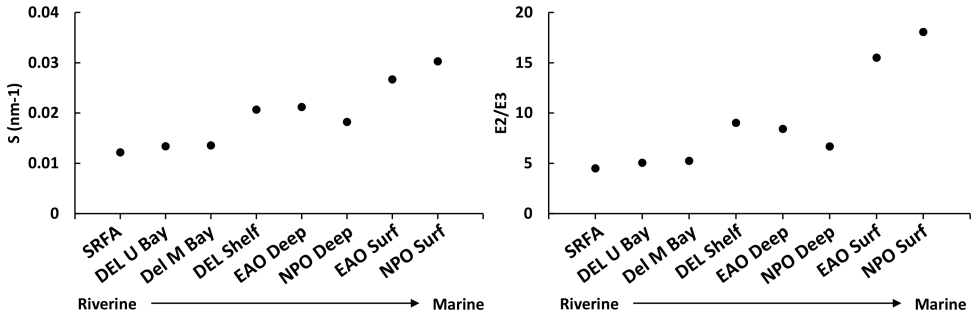


Figure 6.4: Left) Spectral slope from 275-295 nm ($S_{275-295}$) and Right) $E2/E3$ ratio for untreated comparison samples.

Van Krevelen diagrams were made to further visualize the differences and similarities among five of the comparison samples (SRFA, DEL M Bay and Shelf, and NPO Deep and Surf) (Figure 6.5). All samples have a significant number of formulas at higher H/C ratios and are centered around O/C= 0.5, with the highest intensity peaks (largest circles) in the middle of the plot. Some of these formulas make up a common core of species located at the at H/C=∼1.2 and O/C∼0.5, which all of these samples share (Appendix G Figure G.8). Moving from riverine to marine

samples (SRFA > DEL M Bay > DEL Surf > NPO Deep and Surf) there is a clear loss of formula for low H/C ratios (<1) and for high O/C ratios (>0.75).

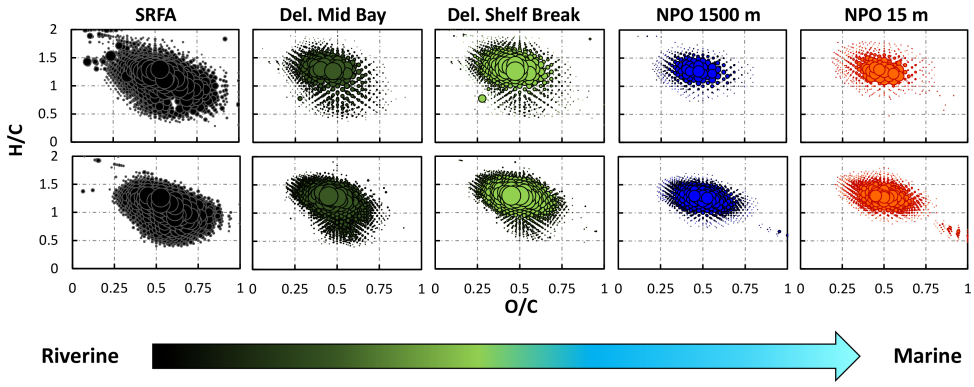


Figure 6.5: Van Krevelen plots of CHO formulae identified in five untreated comparison samples in two mass ranges; Top) 200-400 and Bottom) 400-600 m/z.

Furthermore, to gain better insight into the loss of peaks from riverine to marine samples, Van Krevelen diagrams were created specifically for the three Delaware samples. The loss of CHO formulae are shown from 1) the most inland Delaware sample (DEL U Bay) to the middle bay (DEL M Bay), 2) the middle bay to off shore, the Delaware shelf break (DEL Shelf), and 3) the range from the most inland to the most off shore sample (DEL U Bay to DEL Shelf, respectively) (Figure 6.7). The VanKrevlen diagrams clearly depict a loss of CHO formulae from DEL U Bay to DEL M Bay to DEL Shelf break with the greatest losses for all at masses >400 m/z. The absorption spectra for DEL U Bay and DEL M Bay have a very similar shape with DEL M Bay having slightly less absorption in the visible leading to a slightly higher spectral slope and $E2/E3$ ratio while DEL Shelf has higher absorption in the UV and even less absorption than DEL M Bay in the visible region leading to an

even higher spectral slope and $E2/E3$ ratio as shown in Figure 6.4 above (Delaware absorption comparison provided in Figure 6.6 below). These results suggest the increasing absorption loss in the visible region, from inland to more off shore Delaware samples is correlated with the loss of CHO formulae. As absorption loss increases, the amount of CHO formulas lost also increases, as shown in the VanKrevelen plots below. For DEL U Bay -DEL M Bay there is a small loss of CHO formula, with a greater loss from DEL M Bay- DEL Shelf, and the greatest loss from DEL U Bay - DEL Shelf. These results suggest that there is a correlation between the loss of unsaturated hydrocarbons (low H/C ratio formulae) and the loss of aromaticity (increase in $E2/E3$) with the loss of long wavelength absorption (Figure 6.6 -c&d).

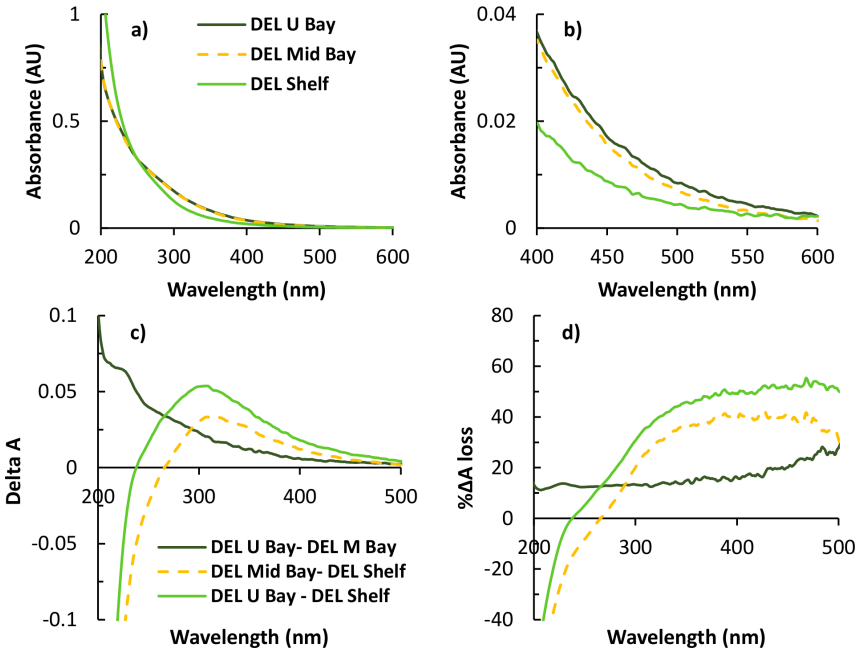


Figure 6.6: a) Absorption for Delaware samples normalized at 250 nm from 200-600 nm; b) Absorption in the visible region 400-600 nm; c) Delta A, change in absorption between samples; d) Percent absorption loss between DEL U Bay and DEL M Bay (dark green), DEL M Bay and DEL Shelf (yellow), DEL U Bay and DEL Shelf (light green).

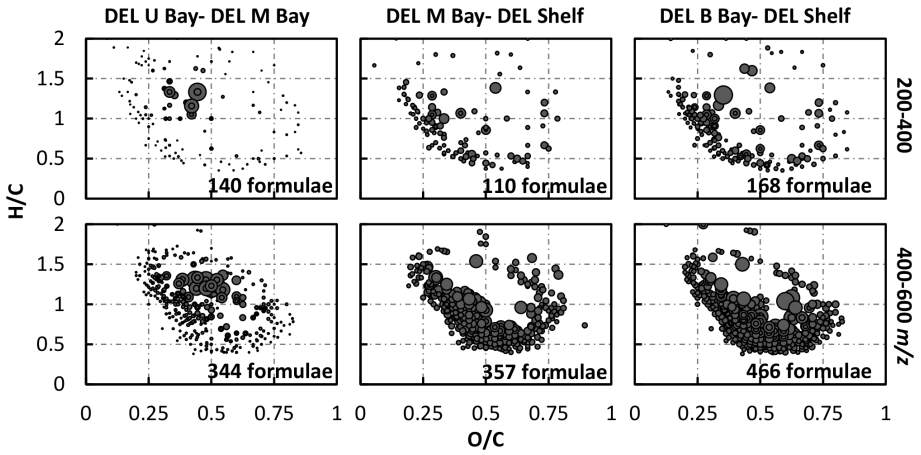


Figure 6.7: Van Krevelen plots comparing the loss of CHO formula form the Delaware upper bay to the Delaware shelf break in three mass ranges (200-400 and 400-600 m/z).

Reduced samples. As seen in previous Chapters and studies [50, 64, 65, 101], reduction of CDOM samples resulted in a loss of absorption at wavelengths > 250 nm and preferential loss at longer wavelengths (SRFA and DEL absorption shown in Figure 6.8; NPO and EAO absorption shown in Figure 6.9), which indicates that all samples have reducible groups (e.g. ketones/aldehydes) that contribute to the long wavelength absorption. The shoulder at $\sim$425 nm in the untreated deep ocean samples is also observed in the reduced samples, suggesting that the species responsible for this shoulder does not undergo reduction. Also, after reduction there was an increase observed in spectral slope and $E2/E3$ ratio due to the preferential loss of visible absorption upon reduction (Figure 6.10). Overall, the ocean samples exhibited far less of an absorption loss than the SRFA and DEL (more riverine) samples (though, to note the ocean samples already had much less absorption in the visible to begin with).

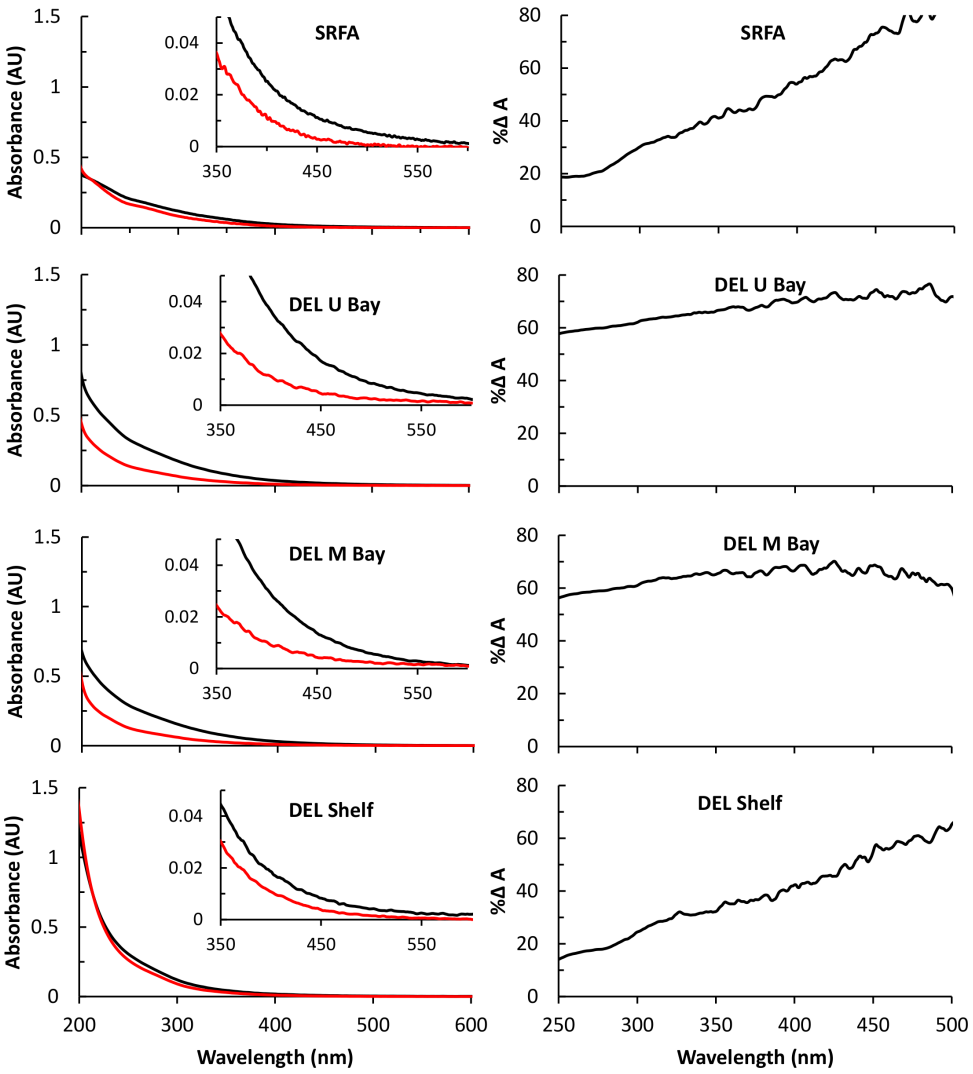


Figure 6.8: Left) Absorption of untreated (black) and 24 hour-reduced (red) samples for SRFA and DEL select transect samples. **Right)** % ΔA, percent loss of absorption, (relative to untreated) following reduction.

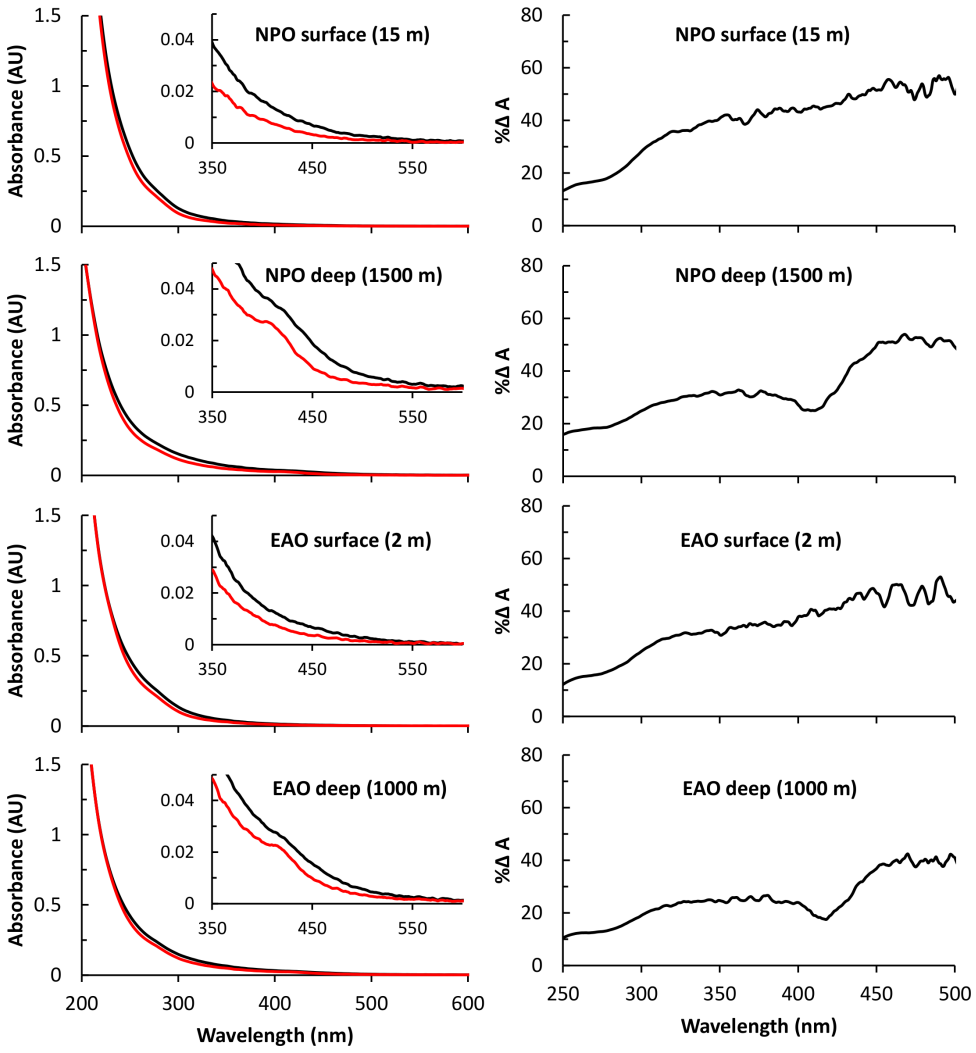


Figure 6.9: Left) Absorption of untreated (black) and 24 hour-reduced (red) samples for select surface and deep NPO and EAO samples. **Right)** % ΔA, percent loss of absorption, (relative to untreated) following reduction.

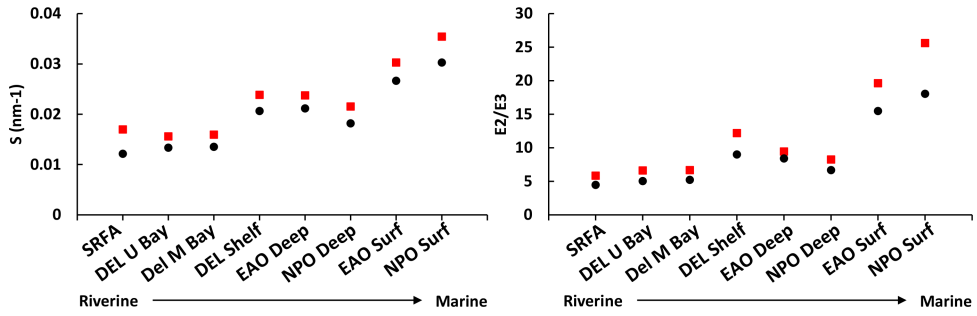


Figure 6.10: Left) Spectral slope from 275-295 nm ($S_{275-295}$) and Right) $E2/E3$ ratio for reduced samples in red and untreated in black for reference.

The mass labeling method using $NaBD_4$ as described in Chapter 3 was used in this study to identify the reduced ketone/aldehyde-containing species in the eight samples chosen for comparison. The untreated samples were used as controls by searching for the unique $M+3n$ (n_{max}=2) in the respective untreated peak lists. The controls resulted in falsely identifying less than 1% of CHO molecular formulae as reduced species (Appendix G Figure G.6, [0,0,2]) which resulted in very low $\%I_{total}$ (sum of $\%I_{1O}$, $\%I_{1B}$, $\%I_{2O}$, and $\%I_{2B}$), indicating that this method generates very few false positives for all of the comparison samples.

Consistent with the trend in absorption, the amount of reduced species decreases dramatically from riverine to marine samples (Figure 6.11). SRFA has the highest amount of reduction ($\%I_{Total}$ ~55%), followed by the DEL U and DEL M Bay ($\%I_{Total}$ ~50% and $\%I_{Total}$ ~40%, respectively), while the marine samples have the least amount of reduction ($\%I_{Total}$ ~10-15%) (Figure 6.11 and Figure 6.12), suggesting that the majority of ketone/aldehyde-containing species contribute significantly to the long wavelength absorption as observed in the terrestrially influenced

samples.

In more riverine like (SRFA, DEL U Bay, and DEL M Bay) samples, 1B (both $M+3$ and $M+6$ but accounting for the intensity of $M+3$ peak) species contribute most to the total reduction, while in more marine like (DEL Shelf, EAO, and NPO), samples, 1O ($M+3$ only) species contribute the most to the total reduction (Figure 6.12). Specifically, for the three Delaware samples, DEL M Bay looks like the cross over point between the highest reduction contribution coming from 1B species to then coming from 1O species (as seen in DEL Shelf). The loss of contribution from 1B species is possibly due to the loss of 2B species in DEL M Bay and DEL Shelf, which in turn decrease 1B species (because they are both from the same original M; therefore 1B species become 1O species due to the loss of the $M+6$ (2B) species). This shift from 1B species to 1O species due to the loss of 2B species as analyzed from DEL U Bay to DEL M Bay and DEL M Bay to DEL Shelf: $\sim$37% of 1O peaks in DEL M Bay were 1B peaks in DEL U Bay and $\sim$24% of 1O peaks in DEL Shelf Bay were 1B peaks in DEL M Bay. This result suggests 2B species are possibly being (photo)degraded into 1O species (this shift of 1B to 1O was also seen in R50x to RED IRR320 from Chapter 5, $\sim$21% of 1O peaks in RED IRR320 were 1B peaks in R50x).

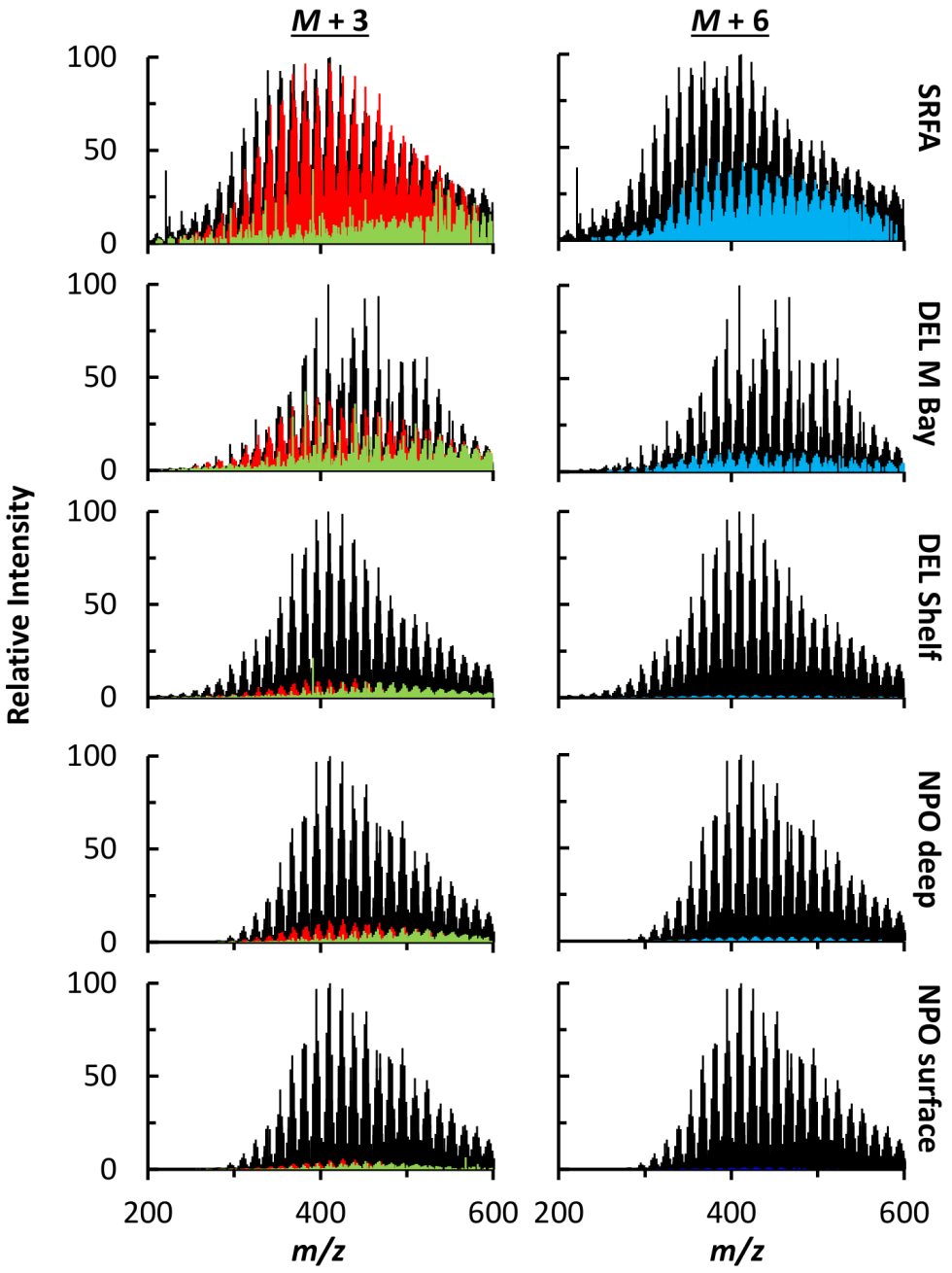


Figure 6.11: Negative ion ESI FT-ICR mass spectra (200-600 m/z) of peaks assigned CHO molecular formulae showing the reduction for eight samples from different aquatic environments. Untreated sample (black), overlaid with respective reduced sample. Intensity of reduced peaks: 1O (green, M+3 only), 1B (red, both M+3 and M+6, showing intensity of M+3 peak), and 2B (blue, both M+3 and M+6, showing intensity of M+6 peak). 2O not shown (less than 2% of the peak intensity for all samples).

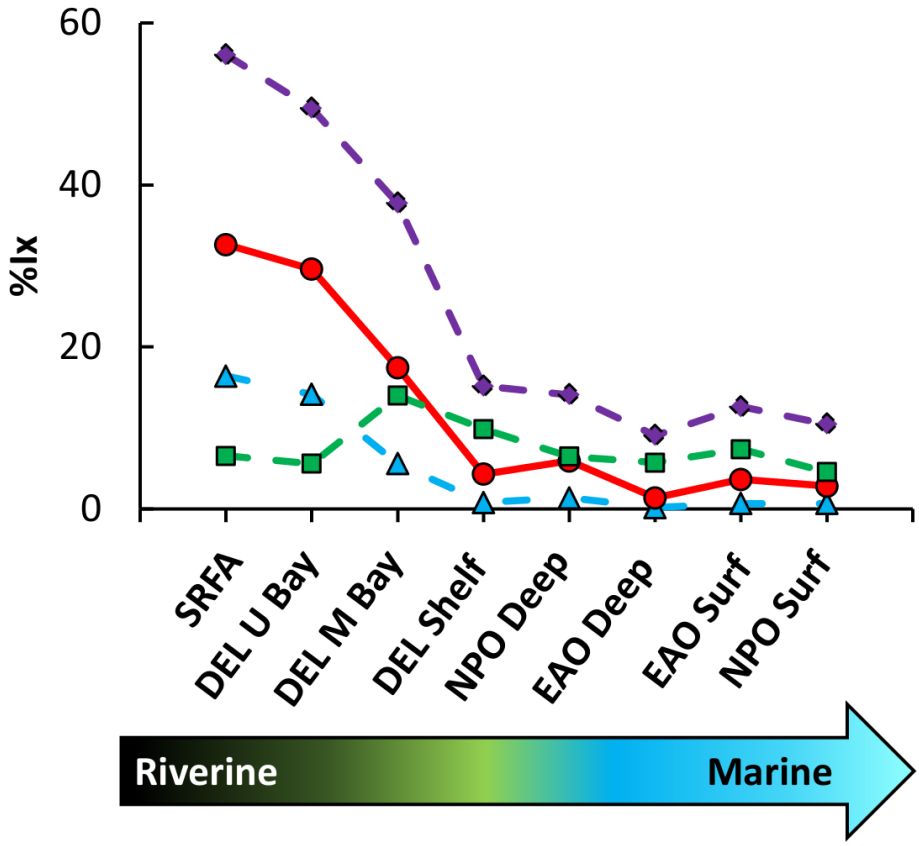


Figure 6.12: $\%I_X$ extent of reduction vs different aquatic environments. $\%I_X$ (intensity-weighted percent of ions across 200-600 m/z). Total percent of reduced intensity (purple, $\%I_{total}$) allocated into groups: $\%I_{1O}$ (green) % of intensity of M+3 only peaks); $\%I_{1B}$ (red) both M+3 and M+6 found, showing % of intensity of M+3 peaks; $\%I_{2B}$ (blue) both M+3 and M+6 found, showing % of intensity of M+6 peaks; $\%I_{2O}$ not shown (% of intensity of M+6 only peaks account for less than 2% of the peak intensity for all samples).

Van Krevlen diagrams were created for all eight reduced comparison samples and divided into low (200-400 m/z, Figure 6.13) and high (400-600 m/z, Figure 6.14) mass regions. The low mass regions have wider distributions than the higher mass regions which are more densely populated towards the center of the plot for both the untreated and reduced samples. The reduced species with the highest intensity

for all of the samples, 1B (red) reside around the center of the plot (H/C$\sim$1.2 and O/C$\sim$0.4), suggesting these species are easily reducible and in all of the aquatic environments. Furthermore, reduced 1O (M+3) species (green) are across all H/C and O/C ratios, while as the samples move to more marine like environments, the amount and intensity of 1B and 2B species decrease and shift to higher H/C ratios. The differences observed in reduction, suggests there is some structural diversity between the DOM material from different aquatic environments; however, there also seems to be a clear transition from terrestrial to marine samples (both untreated and reduced) with the loss of lower H/C formulae and the common core of formulae, above an H/C ratio of 1, that are consistent in all samples (reduced in common formula provided in Appendix G Figure G.9).

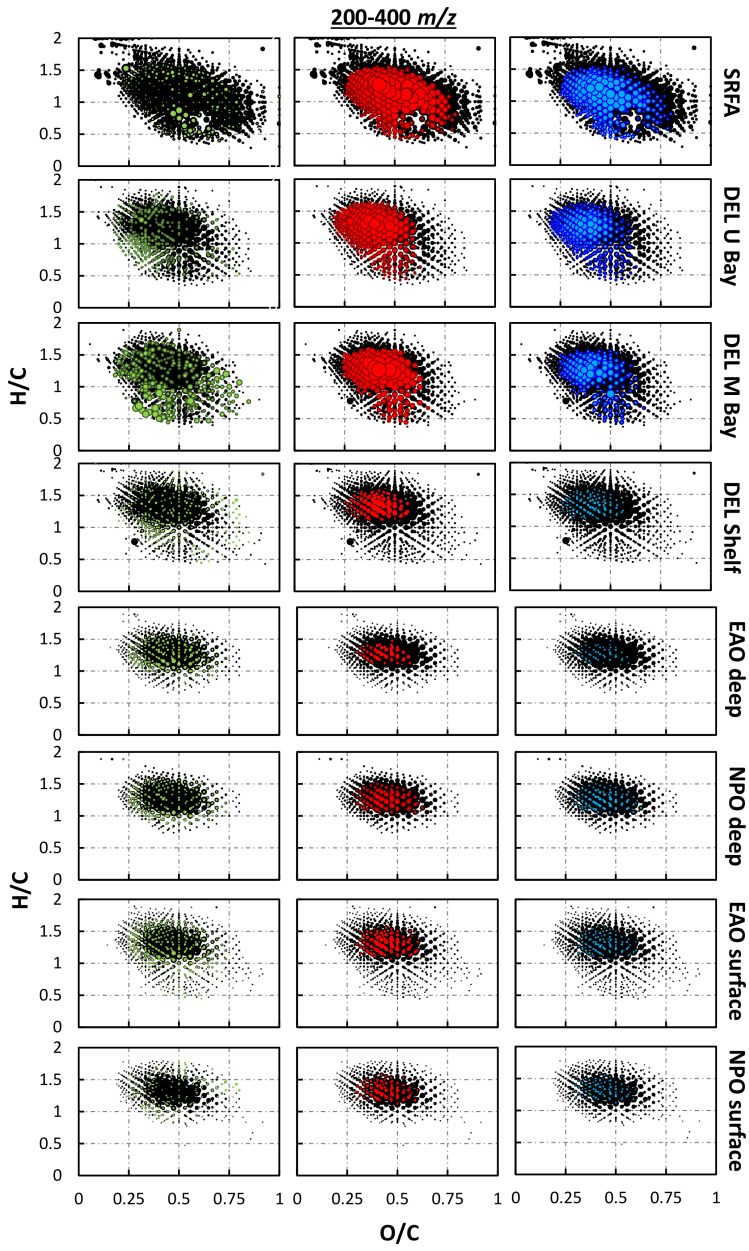


Figure 6.13: Van Krevelen plots of CHO formulae identified in the low (200-400 m/z) mass range showing the reduction for eight samples from different aquatic environments. Untreated sample shown in black, overlaid with the intensity of the respective 24 hour-reduced peaks arising from reduction by borodeuteride. Intensity of reduced peaks color coded: 1O (green, M+3 only),1B (red, both M+3 and M+6, showing intensity of M+3 peak), and 2B (blue, both M+3 and M+6, showing intensity of M+6 peak). 2O not shown (less than 2% of the peak intensity for all samples).

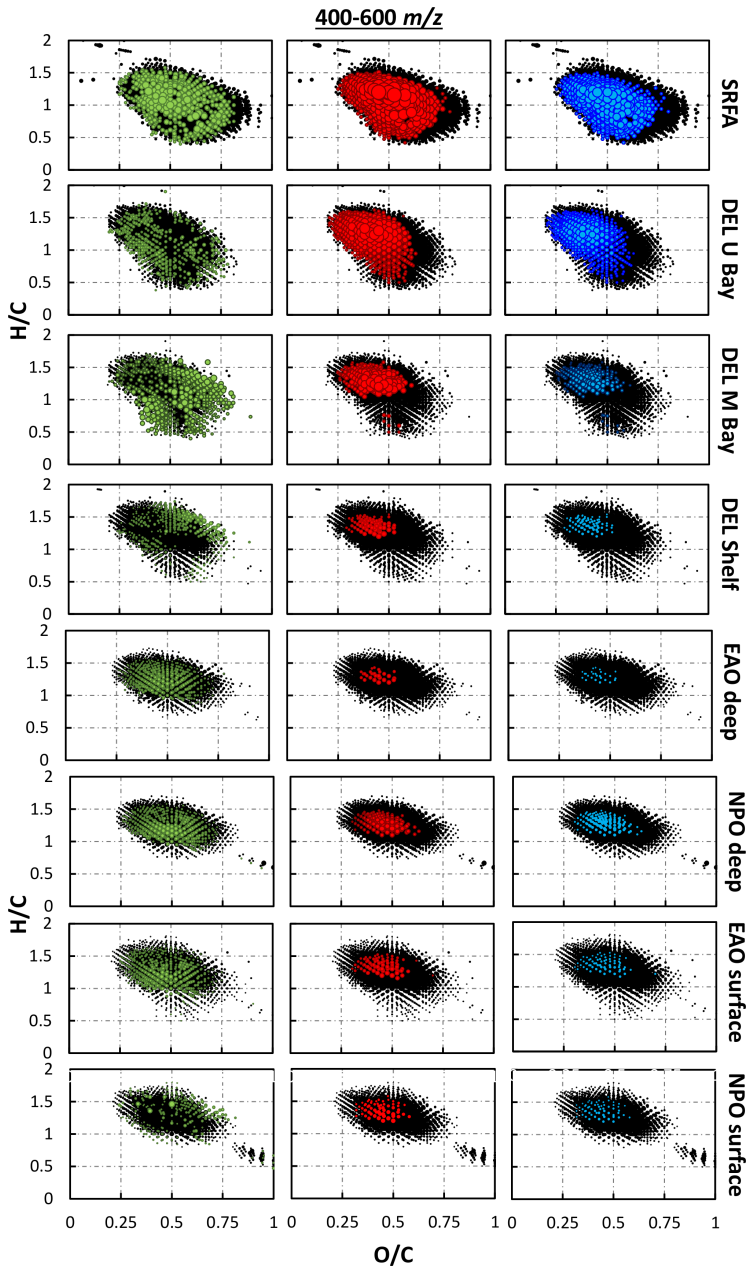


Figure 6.14: Van Krevelen plots of CHO formulae identified in the high (400-600 m/z) mass range showing the reduction for eight samples from different aquatic environments. Untreated sample shown in black, overlaid with the intensity of the respective 24 hour-reduced peaks arising from reduction by borodeuteride. Intensity of reduced peaks color coded: 1O (green, M+3 only), 1B (red, both M+3 and M+6, showing intensity of M+3 peak), and 2B (blue, both M+3 and M+6, showing intensity of M+6 peak). 2O not shown (less than 2% of the peak intensity for all samples).

The percent of integrated absorption loss upon reduction with $NaBD_4$ was correlated to the $\%I_{total}$ calculated from the MS data for three wavelength ranges 250-300 nm, 300-400 nm, and 400-500 nm, respectively (Figure 6.15). In Figure 6.15, the offshore samples cluster together at low absorption loss and low $\%I_{total}$, DEL U Bay and DEL M Bay cluster at high absorption loss and high $\%I_{total}$, while SRFA has low absorption loss and high $\%I_{total}$. When excluding SRFA, there is a correlation within all three wavelength ranges, the strongest from 250-300 and 300-400 nm with an R^2 of 0.92 and 0.87, respectively. The lack of absorption loss vs high reduction calculated from MS for SRFA, suggests that SRFA has a large amount of ketone/aldehyde containing species that are able to be reduced but do not contribute as significantly to the absorption (are not as optically active) as for example, the ketone/aldehyde containing species in DEL U Bay and DEL M Bay (also the difference is possibly due to the different SPE used- XAD-8 resin for SRFA and C_{18} for the extracts). Though, all samples do support the conclusion that reduction of ketone/aldehyde-containing species are primarily responsible for the observed changes in absorption as previously described by the EI model (most likely via participation in chromophore-chromophore interactions).

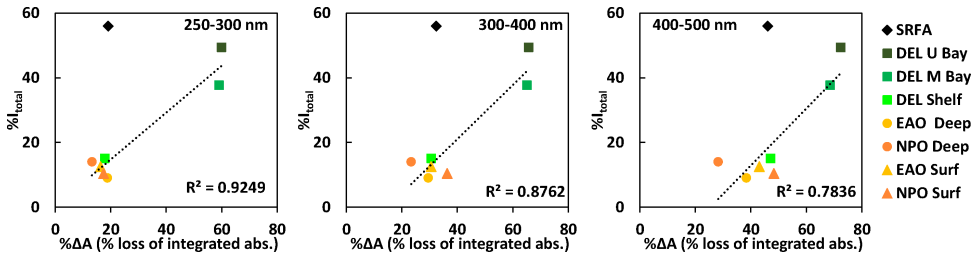


Figure 6.15: Percent of integrated absorption loss upon reduction with $NaBD_4$ for the eight comparison samples (relative to the respective untreated sample, absorption prior to solid-phase (PPL) extraction at pH7) across three wavelength ranges correlated to $\%I_{total}$. 250-300 nm, 300-400 nm, and 400-500 nm. SRFA (diamond), Delaware samples (squares), deep ocean samples (circles), surface ocean samples (triangles). R^2 excludes reduced SRFA sample (R50x).

6.3.4 Similarities Between Samples from Different Aquatic Environments

Heatmaps based, on Bray-Curtis analysis, were created to visualize the similarities in the eight samples. For the Bray-Curtis analysis, the sample size (in this case the number of CHO molecular formula), should be the same for all compared samples; therefore, for untreated samples 1,900 of the highest intensity peaks were analyzed and for reduced samples 800 of the highest intensity reduced peaks were analyzed (Figure 6.16 and 6.17, respectively). The two deep ocean samples (NPO deep and EAO deep) have the most CHO formulae in common (Bray Curtis similarity percent ~90%); the ocean samples in general have the highest amount of CHO formula in common even though they are from different oceans (~81-90%), the ocean samples also share more commonality (~81-87%) with DEL Shelf than

the more riverine samples (SRFA, DEL U Bay, and DEL M Bay). This observation suggests off shore (marine like) samples share a lot of the same formulas even though they are from very different aquatic locales. Furthermore, the more riverine samples, SRFA, DEL U Bay, and DEL M Bay have >65% similarity. Of these samples, DEL U Bay and DEL M Bay are the most similar (∼83%), which is consistent with these samples being relatively close in proximity along the Delaware transect. SRFA shares the most commonality with DEL U Bay (∼71%), and the least commonality with NPO Surf (the commonality for SRFA decreases moving towards more offshore samples; though, SRFA still has ∼56% similarity considering CHO formula with NPO Surf).

There is commonality between all the eight samples, which suggests there may be a subset of formula that are likely to be in nearly all DOM samples that can be detected by ESI MS. As well, Lechtenfeld et al 2014 observed 361 common fomulae in marine DOM samples, which they referred to as the "island of stability" [123]. Though, due to DOM being complex, there are many possible isomers for each molecular formula and thus samples can look more similar to each other than they truly are [124]; therefore, to reveal more about structural details, other methods or chemical tests should be employed.

Figure 6.17 shows the similarity of the 800 highest intensity CHO reduced molecular formulae. Once again, like the untreated samples, the deep ocean samples have the most reduced CHO formulae in common (∼76%) and in general the ocean samples have the most reduced formulae in common (∼73-76%), with the exception of NPO Surf, which has the least in common with all other samples (the very

least with SRFA ∼42%). Using the common peaks in the eight untreated samples and searching for reduced species in SRFA and NPO Surf, it was observed that ∼87% of the reduced peaks (M+3 and M+6) in NPO Surf were also found in reduced SRFA. Futhermore, the Delaware samples have the second highest amount of reduced formulae in common, which once again is consistent with their relatively close proximity, DEL Shelf also shares ∼66-71% of reduced molecular formulae with the ocean samples.

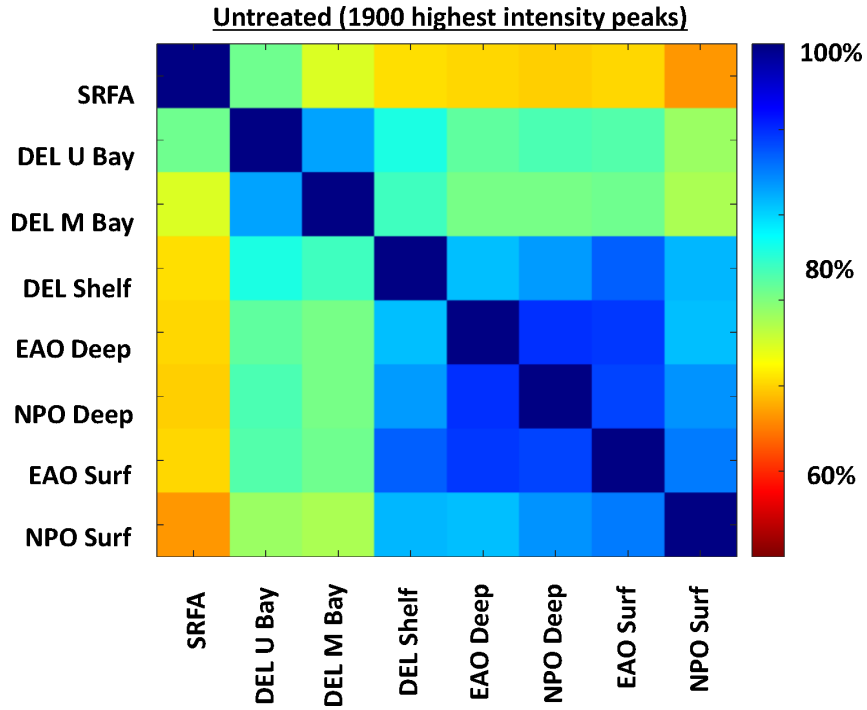


Figure 6.16: Heatmaps showing the similarities (based on Bray-Curtis analysis) of the CHO molecular formula in eight untreated samples from different aquatic locations (1,900 of the highest intensity peaks within 200-600 m/z).

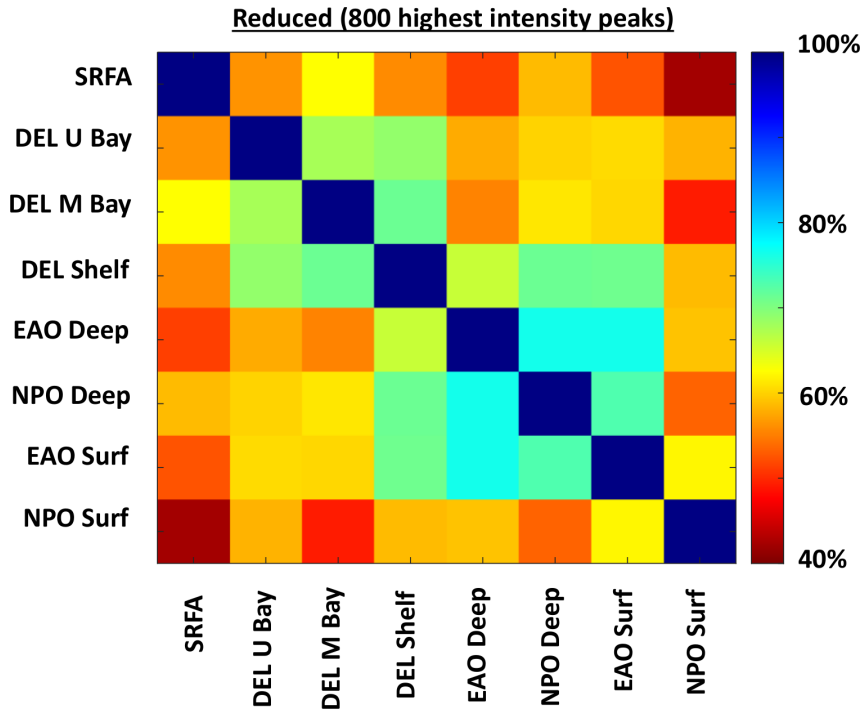


Figure 6.17: Heatmaps showing the similarities (based on Bray-Curtis analysis) of the reduced CHO molecular formula in eight samples from different aquatic locations (800 of the highest intensity reduced peaks from 200-600 m/z).

6.4 Conclusions

This chapter focused on the characterization of the optical properties and mass spectrometric properties, through the use of deuterium labeling by $NaBD_4$ and ESI FT-ICR-MS, of eight DOM samples from different aquatic locations. An increase in spectral slope and $E2/E3$ ratio, for untreated samples, was observed moving from riverine to marine samples suggesting that marine samples are less aromatic than riverine samples, which correlated well with loss of lower H/C formulae (unsaturated hydrocarbons) seen in the VanKrevelen diagrams. These results suggest, as seen

specifically for the Delaware transect samples, that the loss of lower H/C formulae is correlated to the loss of longer wavelength absorption.

Furthermore, riverine (DEL U bay and DEL M Bay) samples showed the greatest loss in absorption after reduction compared to the marine samples, suggesting that the riverine samples have a higher amount of ketone/aldehyde-containing species. The results from the mass labeling method are in agreement, as there was a higher percentage of reduced species observed in riverine samples. Thus, supporting the theory that ketones/aldehyde contribute to the long wavelength absorption (chromophore-chromophore interactions) interactions in the EI model.

The use of the molecular formulae assignments and VanKrevelen plots along with the mass labeling method has helped distinguish similarities (especially at H/C ratios >1) and differences (below at H/C ratios <1) within the samples as well as highlighted the differences in the reduced species, giving insights into the structure of DOM from different aquatic environments. Heatmaps constructed from Bray-Curtis analysis showed the commonality between CHO molecular formula in both untreated and reduced samples. High commonality in the H/C>1 in the reduced samples across environments suggests a removal of H/C<1 (from the terrestrial material) by chemical/microbial processes (one such being photochemistry) [106] with the remnants remaining relatively stable (and of higher spectral slope).

Chapter 7: Conclusions and Future Work

7.1 Conclusions

This work describes and applies a mass labeling method using ultrahigh resolution mass spectrometry and deuterium labeling to determine the contribution of ketone/ aldehyde-containing species to the composition and optical properties of dissolved organic matter. The contribution of ketone/aldehyde-containing species was investigated by reducing DOM samples with $NaBD_4$ and observing the changes due to reduction through mass spectrometry and UV-Vis and fluorescence spectroscopy. The mass labeling method assigns molecular formulae and identifies peaks which have been reduced by using the assigned molecular formulae from the untreated samples and searching the reduced spectrum for $M+3$ and $M+6$ species. This algorithm for identifying reduced species was tested for internal consistency using two additional methods, both of which resulted in consistently identified reduced species in the mass spectra (Chapter 3).

Through the use of a $NaBD_4$ titration study of a reference material SRFA, it was demonstrated that different compounds are reduced under low and high amounts of $NaBD_4$, indicating tthat these reduced species have multiple reducible moieties with varying reactivity and accessibility (Chapter 4).The effects of irradiation on

SRFA was also examined (Chapter 5), which revealed insights into the structural components in SRFA that were lost or decreased due to irradiation, as well as how the effects of irradiation can provide additional insight concerning the lack of mass spectral components in open ocean samples that are observed in samples of terrestrial origin.

A comparison study analyzing eight DOM samples from diverse aquatic environments was conducted to investigate the differences and similarities between DOM/CDOM found in these environments (Chapter 6). The absorption spectra of samples ranging from riverine to coastal to open ocean samples share many of the same optical properties as has been observed previously [10,50] and respond similarly to pH and borohydride reduction, suggesting a possible terrestrial-like component in marine samples.

The differences observed between riverine and marine samples could in part be due to photodegradation, as shown by the irradiation study and the commonality seen between the riverine and marine samples. This idea is supported by both the optical data and the mass spectrometry as evidenced by the common core of peaks found for both untreated and reduced riverine and marine samples, suggesting a pool of DOM that is not easily degraded (photochemically or photochemically and biologically). This method has proven useful in relating the changes in the optical properties upon the reduction by sodium borodeuteride to the changes observed by mass spectrometry to reveal information on the composition as well as source and structure of DOM from different aquatic environments.

7.2 Future Work

The application of the mass labeling method can be further implemented to characterize and compare DOM from even more (diverse) aquatic environments. In addition, more efforts to quantitatively link ketone/aldehyde-containing species observed by this method to the bulk optical properties can be performed. Combining methods, for example using size exclusion chromatography or liquid chromatography with UV-Vis and MS, could aid in the identification and characterization of subsets of species responsible for long wavelength absorption in the visible.

While this method was tested and applied using ESI, as mentioned in Chapter 1 other soft ionization techniques such as APPI or APCI could be used as complementary techniques to ESI. These other ionization techniques can provide additional insights into species that are not as readily ionized by ESI and offer more information on the composition/structure of compounds with in DOM.

Chapter 5 showed how irradiation combined with MS revealed the effects of photodegradation on SRFA; conducting more irradiation studies examining the wavelength dependence on a larger variety of samples could provide better information on the photochemical reactivity of DOM from differing aquatic environments, which in turn could provide a better understanding of the relationship between riverine and marine DOM.

Appendix A: Internal calibrants and extraction blank

Table A.1: *m/z* and corresponding molecular formulae (of [M-H]- ions) used for post-acquisition internal calibration of the mass spectra.

m/z	**mol. form**	***m/z***	**mol. form**
201.0405	$C_8H_9O_6$	475.0882	$C_{22}H_{19}O_{12}$
225.0041	$C_9H_5O_7$	501.0675	$C_{23}H_{17}O_{13}$
251.0197	$C_{11}H_7O_7$	525.0675	$C_{25}H_{17}O_{13}$
275.0561	$C_{14}H_{11}O_6$	551.0831	$C_{27}H_{19}O_{13}$
301.0354	$C_{15}H_9O_7$	575.1042	$C_{26}H_{23}O_{15}$
325.0929	$C_{15}H_{17}O_8$	601.0835	$C_{27}H_{21}O_{16}$
351.1085	$C_{17}H_{19}O_8$	625.0835	$C_{29}H_{21}O_{16}$
375.0722	$C_{18}H_{15}O_9$	651.0992	$C_{31}H_{23}O_{16}$
401.0514	$C_{19}H_{13}O_{10}$	675.0992	$C_{33}H_{23}O_{16}$
425.1089	$C_{19}H_{21}O_{11}$	701.0996	$C_{31}H_{25}O_{19}$
451.1246	$C_{21}H_{23}O_{11}$		

Table A.2: *m/z* and tentative assignments of peaks present in the extraction blank which were removed from mass spectra

m/z	Identification	*m/z*	Identification
255.23295	unknown (monoisotopic)	339.19991	sulfonate (monoisoptopic)
265.14792	unknown (monoisotopic)	345.28301	unknown
266.15126	unknown (+ ^{13}C- ^{12}C)	359.29866	unknown
283.26424	unknown (monoisotopic)	360.30202	unknown (monoisotopic)
284.26760	unknown (+ ^{13}C- ^{12}C)	391.28850	unknown
293.17921	unknown (monoisotopic)	421.22653	unknown (monoisotopic)
309.10134	unknown (monoisotopic)	422.22991	unknown (+ ^{13}C- ^{12}C)
309.10141	unknown (monoisotopic)	467.15567	unknown
311.16860	sulfonate (monoisoptopic)	473.39998	unknown (monoisotopic)
325.18426	sulfonate (monoisoptopic)	529.46259	unknown (monoisotopic)
331.26737	unknown	530.46595	unknown (+ ^{13}C- ^{12}C)
333.24699	sulfonate (monoisoptopic)		

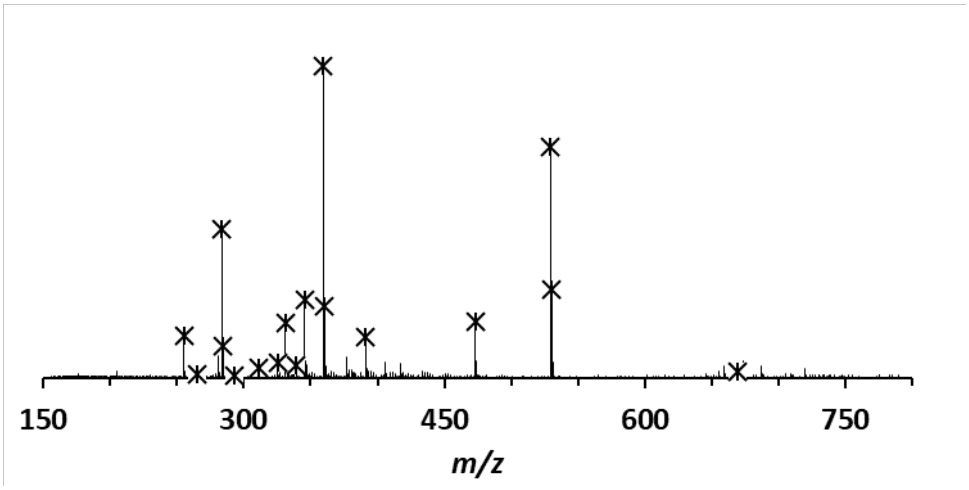


Figure A.1: Negative ion ESI FT-ICR mass spectrum of the extraction blank. Asterisks (x) denote *m/z* peaks which were excluded from sample peak lists.

Appendix B: Isobaric Species

B.1 Isobaric species: reduction with low and high amounts of $NaBD_4$

Most (if not all) m/z peaks in an FT-ICR mass spectrum of DOM have multiple isomers [52, 126]. The chemical reduction by $NaBD_4$ can affect these isomers differently depending on the reactivity as well as accessibility of the reducible group. This was shown when the original M ion was identified as having one *and* two reducible moieties (referred to as 1B and 2B in 3.3). The original peaks at M could be 1) a set of species each with one or two reducible species or 2) an individual species with two reducible groups that was only partially reduced. A hypothetical example of this is shown in Figure B.1.

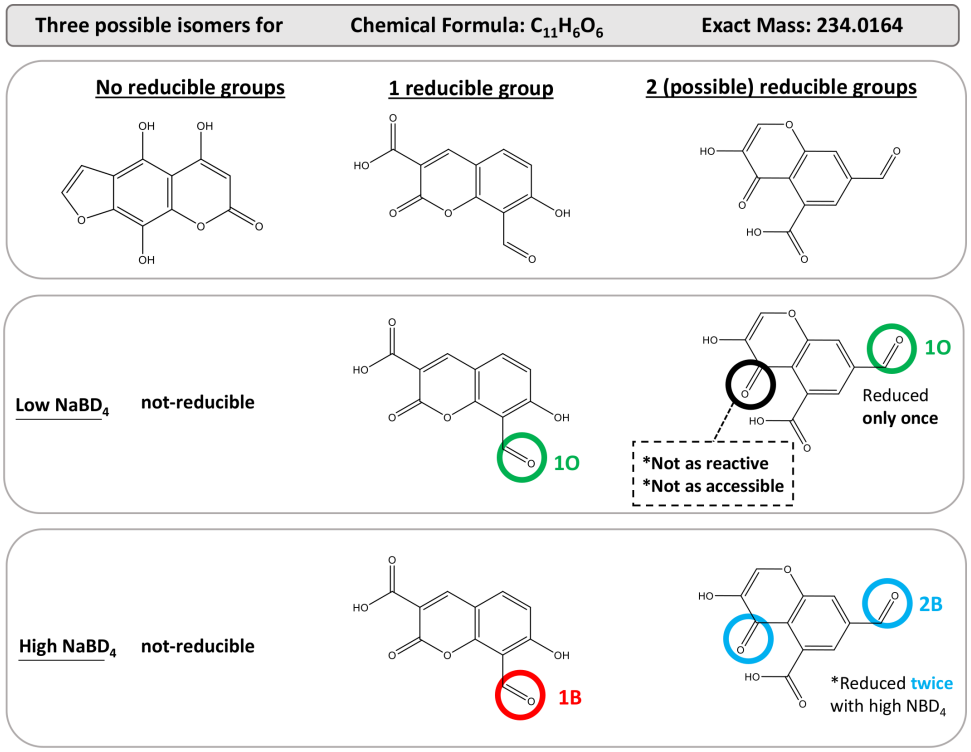


Figure B.1: Potential reductions of possible isomers for a hypothetical DOM molecule. 1O- singly reduced only (green); 2O- doubly reduced only (black); 1B and 2B- singly and doubly reduced (red and blue, respectively). For further descriptions of 1O, 1B, 2O, and 2B, see section 3.3.

Appendix C: Internal consistency

C.1 Calculating %I with three Methods A, B, and C

Internal consistency: To validate our approach to identify deuterium-containing molecular formulae in reduced samples, a two additional methods of calculating the amount of reduction were considered. With Method A (described fully in Chapter 3) molecular formulae are assigned for the untreated sample and reduced peaks are then searched for in reduced samples from the assigned molecular formulae (of the untreated sample).

Method B is a similar method but instead of searching for reduced peaks from assigned formulae in the untreated sample, this method searched for reduced peaks using the untreated and reduced peak list. Additions of M+3.021927 and M+6.043854 were added to the m/z in untreated peak list (UNT_M+3n) which was then compared against the reduced peak list. The M+3.021927 and M+6.043854 peaks found in the reduced peak list were considered reduced.

Method C, like in Method A, molecular formulae were used to calculate reduction, but only the reduced sample was considered and the molecular formulae containing deuterium were considered reduced.

The $\%I_{TOTAL}$ for methods A and C, considering only CHO assignments, were

<1% different and both methods A and B were <5 % different (within 200-600 m/z-Table C.1 and Methods B and C-Figure C.1), showing that all methods consistently identify reduced species in mass spectra.

Table C.1: %I calculated by three methods (from 200-600 m/z).

	Method A[a]			Method B[b]			Method C[a]		
	$M+3$	$M+6$	Total	$M+3$	$M+6$	Total	$M+3$	$M+6$	Total
R3	28	2	29	24	4	28	28	2	30
R11	30	4	34	27	8	32	29	5	34
R23xA	36	8	44	33	9	41	35	9	44
R50x	39	17	56	35	17	52	39	18	57

[a]Formulae (CHO only formula assignments); [b]Peaks

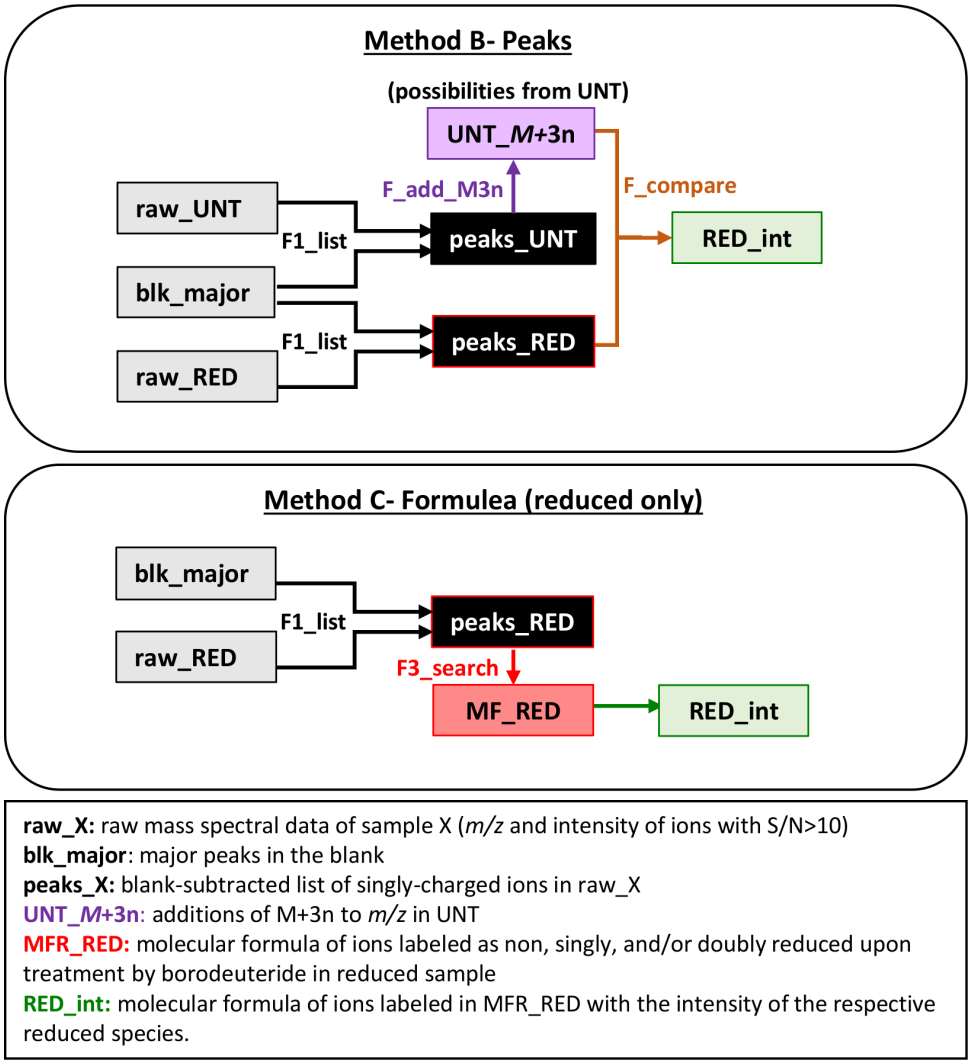


Figure C.1: Illustration of data variables (boxes) and functions (arrows) used for the analysis of mass spectrometric data when testing the internal consistency of the methods (performed in MATLAB) for untreated sample (UNT) and reduced sample (RED).

Appendix D: Methods for Optical Measurements

D.1 Apparatus for Optical Measurements

UV-visible absorbance spectra were collected on a Shimadzu UVPC 2401 spectrophotometer. Fluorescence spectra (three dimensional excitation emission matrices, EEMs) were collected with a Fluoro-Max 4 luminescence spectrometer. For pH measurements a Thermoscientific micro pH electrode coupled to an Orion 4 115 Star pH ISE bench top meter was used.

D.2 Optical Measurements

Prior to analyses, all samples were filtered through 0.2 μm nylon filter to minimize interference (e.g. scatter) from any undissolved particulate matter. All optical measurements were carried out in a 1-cm quartz cuvette. Absorbance spectra were measured against air from 190-800 nm. The absorption spectra of a blank (in Milli-Q water) was recorded and subtracted from all sample spectra. For each absorption spectrum, slight baseline offsets were corrected by subtracting average absorption over the 650-700 nm range [10]. To observe the absorbance changes due to increasing amounts of $NaBD_4$ the percent change of the integrated absorbance

(%ΔA) relative to that of the untreated sample was calculated:

$$\%\Delta A = 100 \times \frac{A_{UNT} - A_{RX}}{A_{UNT}} \tag{D.1a}$$

where A_{UNT} and A_{Rx} are the integrated absorption of the untreated and reduced samples, respectively, over defined wavelength intervals.

Fluorescence spectra were acquired for excitation wavelengths from 240-600 nm at 10 nm wavelength intervals; with the emission from 290-700 nm every 1 nm. Excitation and emission monochromator band passes were set at 4 nm, and spectra were corrected for the instrument response using factors supplied by the manufacturer. Each emission spectrum was blank subtracted, smoothed (with a five-point moving average), and scatter peaks (Rayleigh and Raman) were removed using an excision and interpolation procedure based on the method described by Zepp et al [127] (in MATLAB-scripts in Appendix I).

Emission (Excitation = 350 nm) spectra were acquired for solutions of 10 ppm and 1 ppb quinine sulfate in 0.1 N H_2SO_4, respectively and employed to convert the EEMs fluorescence intensity units into relative quinine sulfate units. All corrected emission intensities were converted to quinine sulfate units (QSU), where 1 QSU is equal the emission intensity of a 1 ppb quinine sulfate solution.

To observe changes in emission intensity due to increasing amounts of $NaBD_4$ the percent change of total emission intensity (from 290-700 nm) relative to that of

the untreated sample was calculated:

$$\%\Delta F = 100 \times \frac{F_{R-X} - F_{UNT}}{F_{UNT}} \tag{D.2a}$$

where F_{UNT} and $\mathrm{F}_R - X$ are the integrated emission of the untreated and reduced samples, respectively, over 290-700 nm.

Appendix E: Supplemental Information for Chapter 4

E.0.1 Comparison of pre and post-extracted SRFA (untreated and reduced)

The absorbance of the post extracted material (Post) is very similar in spectral shape to that of the absorbance of the material prior to extraction (Pre) (i.e. whole material). In support, the Post/Pre is wavelength independent indicating that the majority of chromophoric material in SRFA was successfully retained by the PPL cartridge. Most differences observed are most likely due to a dilution effect (as shown by the normalization of the post to pre at 300 nm) rather than compositional differences between whole and extracted material (Figure E.1. The normalization of post to pre material was calculated as shown by the equations below:

$$\frac{A_{Pre300nm}}{A_{Post300nm}} = NormaliztionFactor_{at300nm} \tag{E.1}$$

$$A_{Post} \times NormaliztionFactor_{at300nm} = Absorbance_{PostNormalized} \tag{E.2}$$

where $A_{Pre300nm}$ is the absorbance of the material at 300 nm prior to extraction, $A_{Post300nm}$ is the material after extraction at 300 nm, A_{Post} is the absorbance at every one nm from 250-500 nm.

This approach was also used to normalize the fluorescence data for %ΔF , calculated as shown by the equations below:

$$\frac{PreF_{ex300,em450}}{PostF_{ex300,em450}} = Normaliztionfactor_{Emission} \tag{E.3}$$

$$PostF_{ex300,em450} \times Normaliztionfactor_{Emission} = Emisson_{PostNormalized} \tag{E.4}$$

where $PreF_{ex300,em450}$ is the emission of the material at excitation 300 nm and emission 450 nm prior to extraction, $PostF_{ex300,em450}$ is the emission of the material at excitation 300 nm and emission 450 nm after extraction, PostF is the emission excitations 240-600 nm at every one nm from 290-700 nm.

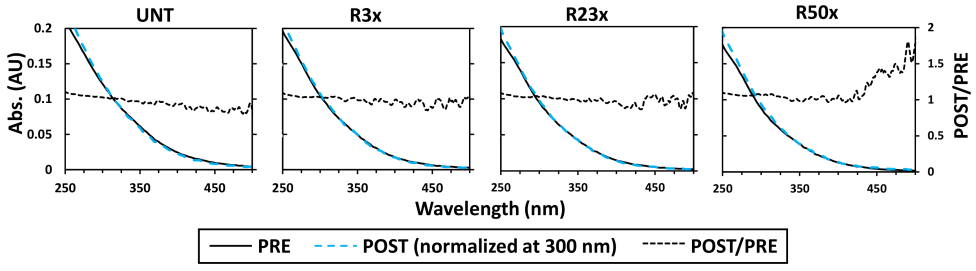


Figure E.1: Absorption spectra at pH 7 before and after solid-phase (PPL) extraction. **Left-axis)** Prior to (Pre) extraction (solid black line), following (Post) PPL extraction absorption of the extracted material normalized to its pre-absorption at 300 nm to account for dilution (dashed blue line). **Right-axis)** Absorption of the material following reduction divided by its absorption before extraction (dashed black line).

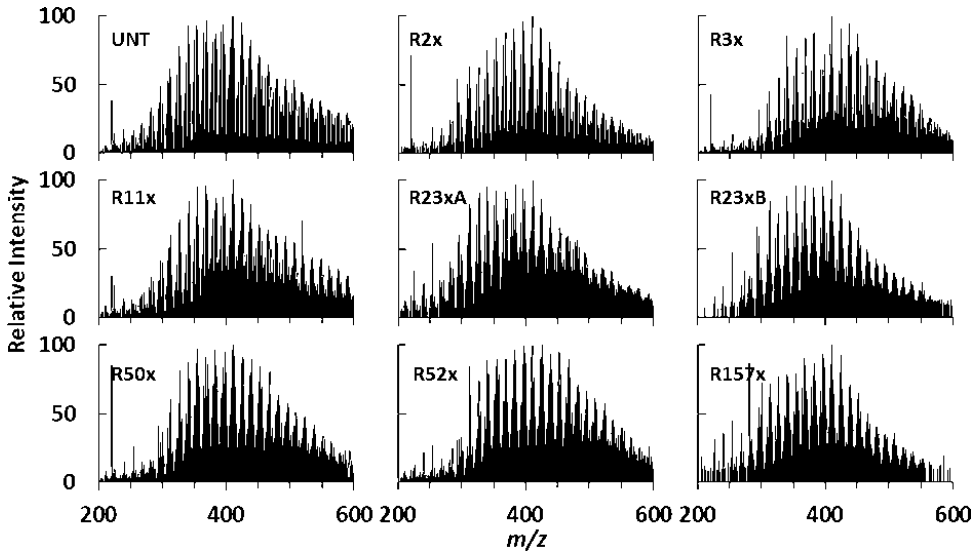


Figure E.2: Negative ion ESI FT-ICR mass spectra of all samples with CHO-only assigned molecular formulae peaks 200-600m/z

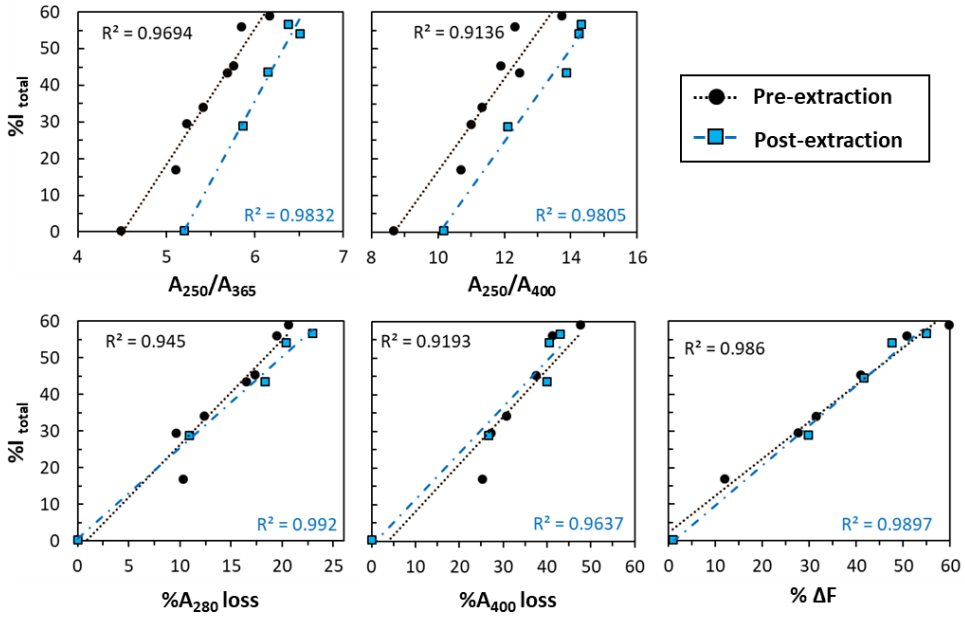


Figure E.3: Optical properties at pH7 correlated to%I_{total} (200-600 *m/z)*. Optical properties prior to solid-phase (PPL) extraction (excluding R157) (black circles) and after solid phase extraction (data only collected for UNT, R3x, R23x, R50x, R52x) (blue squares).

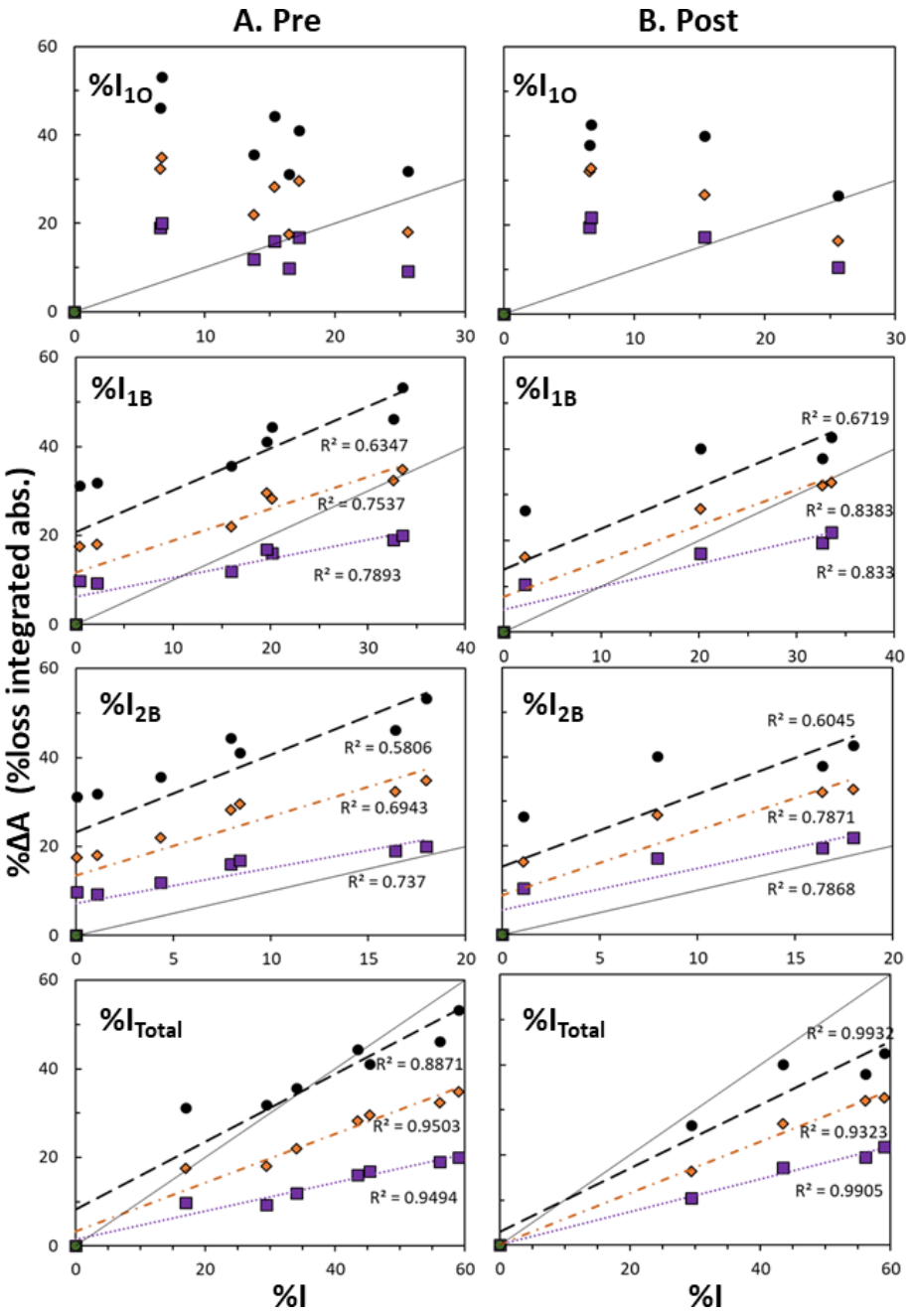


Figure E.4: Percent of integrated absorbance loss upon reduction with $NaBD_4$ relative to untreated (UNT) sample (pH7) across three wavelength ranges correlated to $\%I_{1O}$, $\%I_{1B}$, $\%I_{2B}$, and $\%I_{total}$. 250-300 nm (purple squares), 300-400 nm (orange diamonds), and 400-500 nm (black circles). **A)** Pre: absorbance prior to solid-phase (PPL) extraction (excluding R157) and **B)** Post: absorbance after solid-phase (PPL) (data only collected for UNT, R3x, R23xA, R50x, and R52x) ($\%I_{2O}$ not shown less than 1% of the peak intensity for all samples).

Appendix F: Supplemental Information for Chapter 5

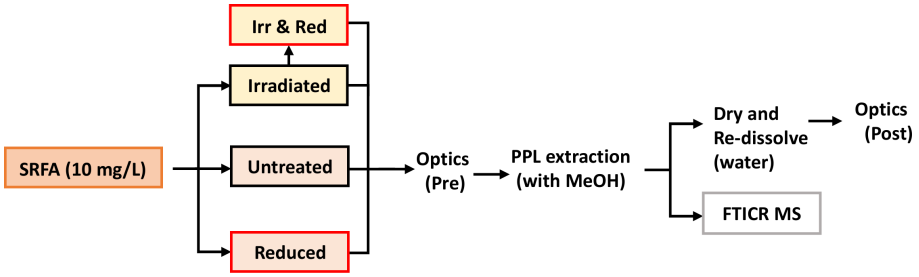


Figure F.1: Block diagram of the sample preparation of untreated, irradiated and reduced SRFA.

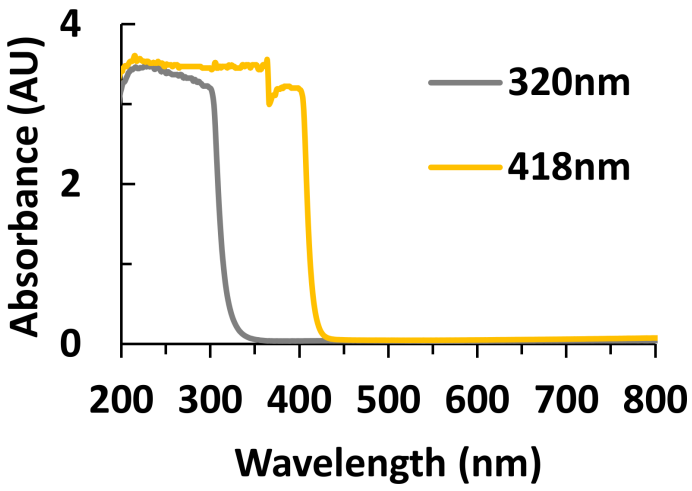


Figure F.2: Absorbance of 418 nm (yellow) and 320 nm (gray) long pass filters used for irradiation studies.

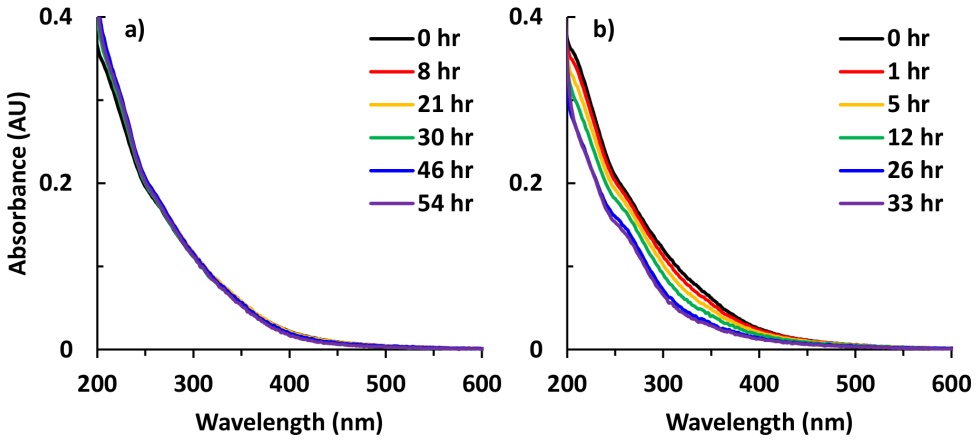


Figure F.3: Absorbance over time for a) 418 nm and b) 320 nm long pass filters.

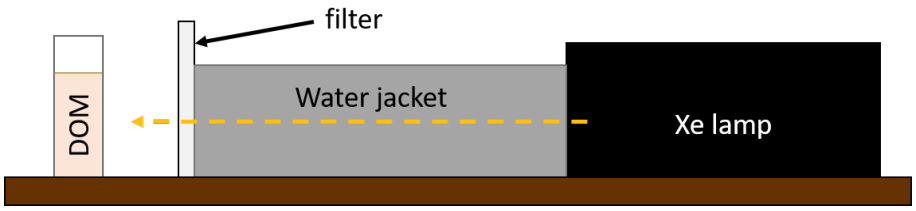


Figure F.4: Irradiation setup. Light source 300 W xenon lamp; 20 cm water jacket was placed between the lamp and the cuvette to remove infrared radiation; long pass filter placed after the water jacket in front of the 1-cm cuvette.

F.1 Comparison of pre and post-extracted SRFA (untreated, irradiated, and reduced)

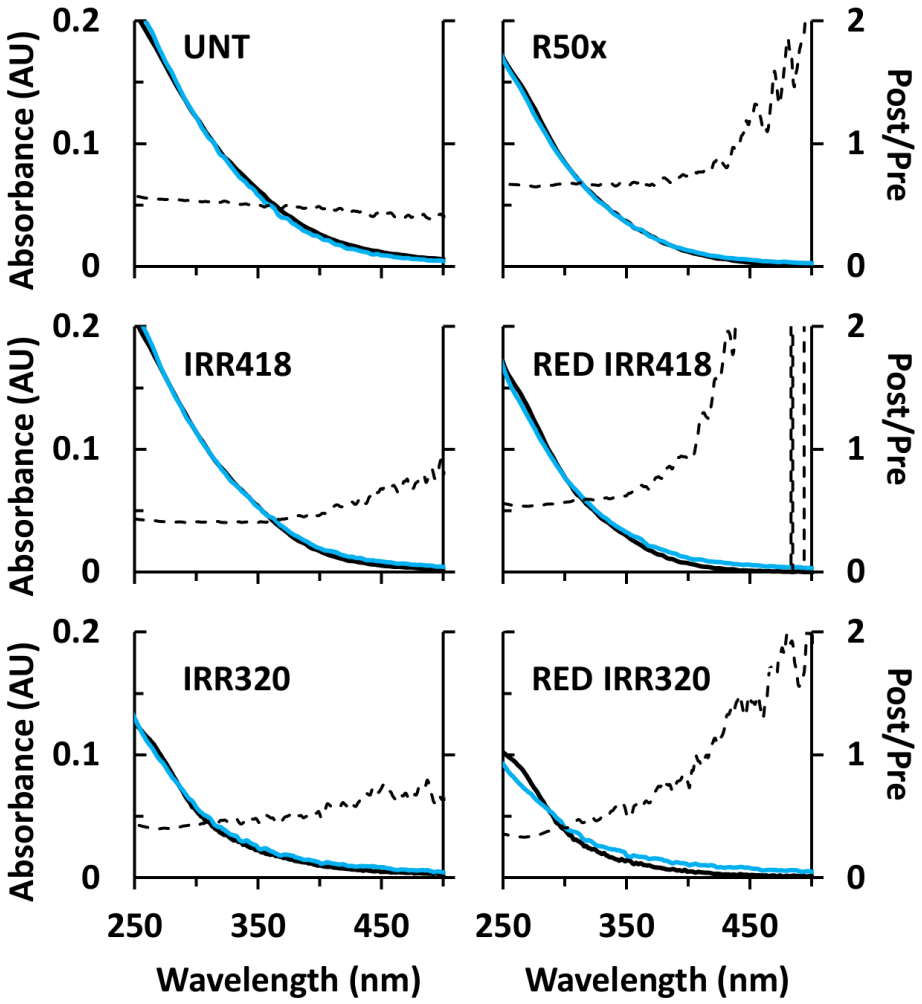


Figure F.5: Absorption spectra before and after solid-phase (PPL) extraction (pH7). Left-axis) Prior to extraction (solid black line), following PPL extraction (dashed blue line), and absorption of the extracted material normalized to its respective pre-absorption at 300 nm to account for dilution (solid blue line). Right-axis) Absorption of the material following extraction divided by its absorption before extraction (dashed black line).

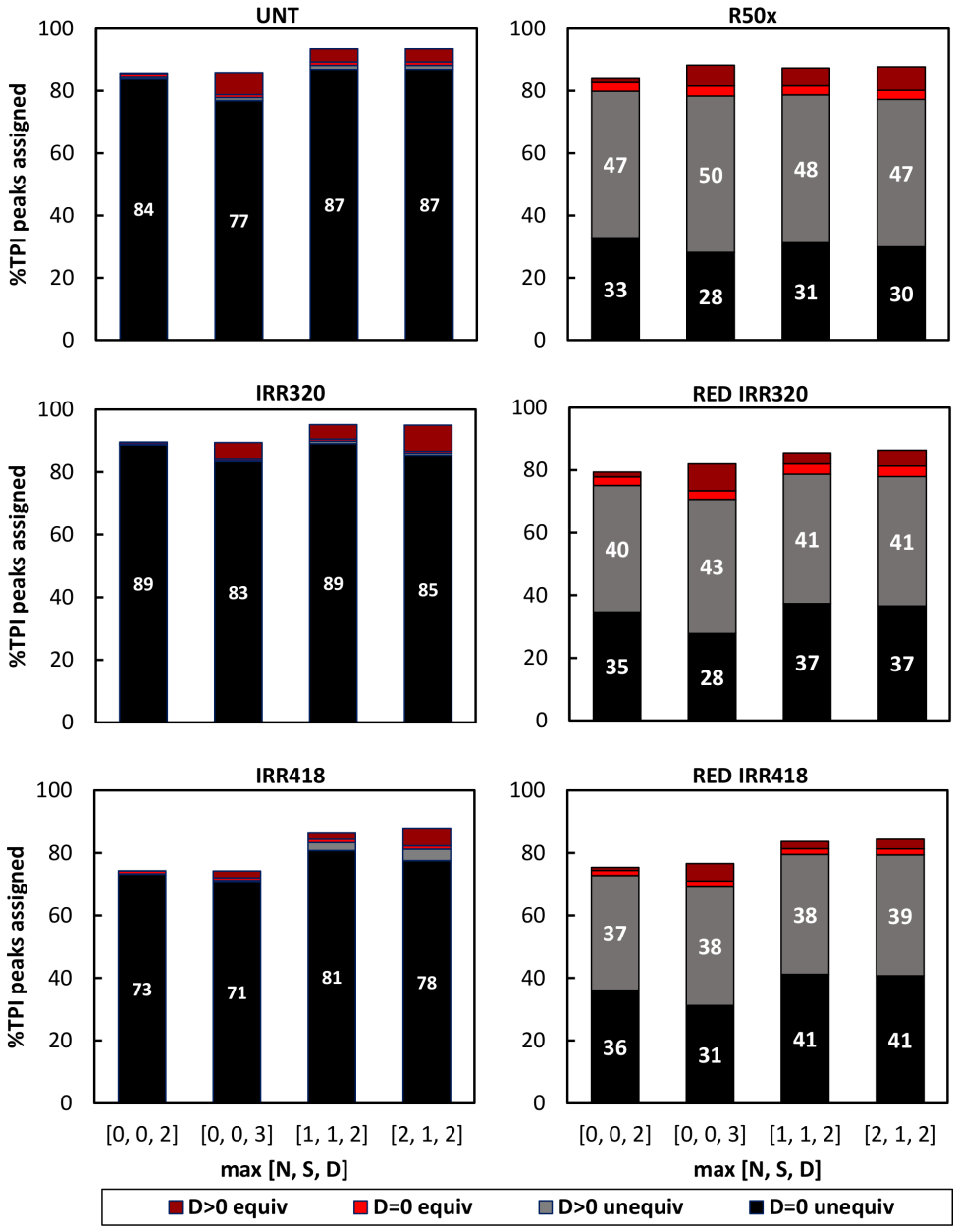


Figure F.6: Bar graph of %TPI of untreated, irradiated, and reduced samples assigned equivocal/unequivocal and D-containing/non-D-containing molecular formulas for various allowances of N, S, D (0,0,2; 0,0,3; 1,1,2; and 2,1,2, respectively) from 200-700 m/z.

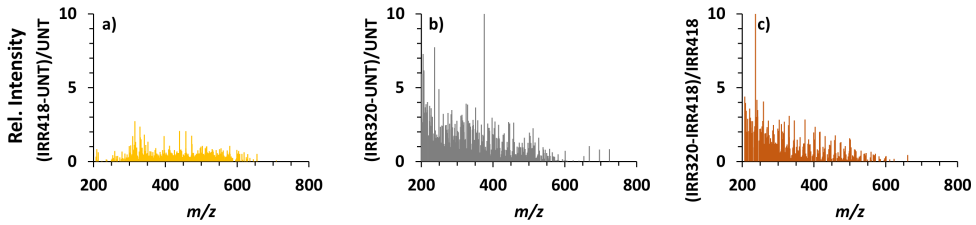


Figure F.7: Increase in peak intensity after irradiation; a&b) from the irradiated to unteated sample; c) from IRR320 to IRR418.

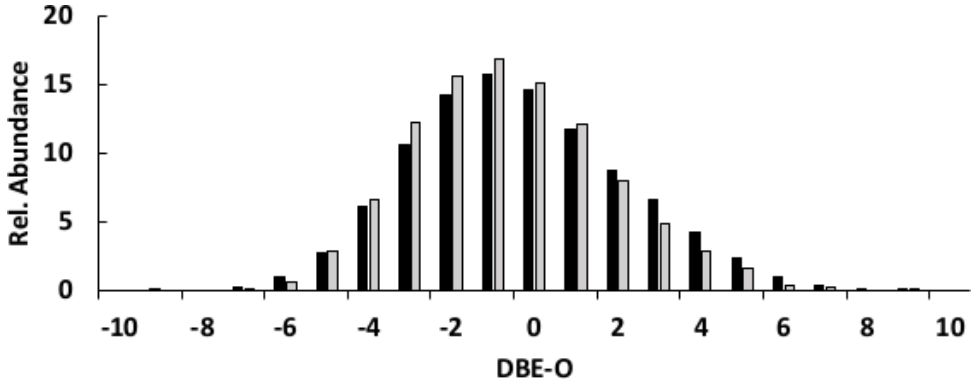


Figure F.8: Realative abundance of DBE-O of untreated and IRR320. Untreated (black); IRR320 (gray).

Table F.1: Percent absorbance loss due to irridiation and reduction.

λ range	Irradation		Reduction			Irr. + Reduction	
	IRR418	IRR320	R50x	IRR418	IRR320	IRR418	IRR320
250-300	3	43	22	22	20	25	54
300-400	14	61	39	42	35	50	75
250-500	11	52	32	32	27	40	65

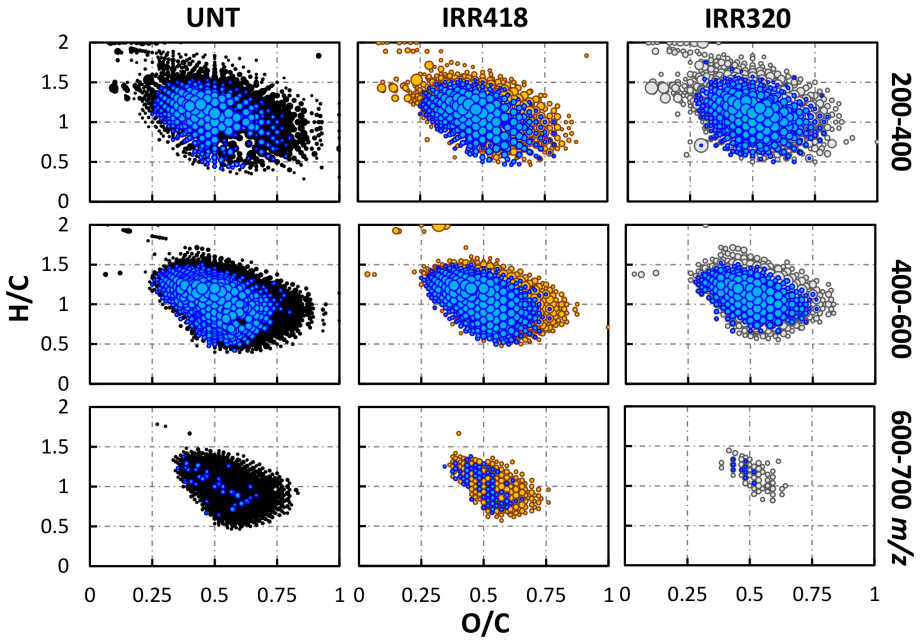


Figure F.9: Van Krevelen plots of 2B CHO formulae identified in untreated, irradiated, reduced, and irradiated and reduced SRFA (UNT, IRR418, IRR320, R50x, RED IRR418, and RED IRR320, respectively) in three mass ranges (200-400, 400-600, and 600-700 m/z). UNT (black), IRR418 (yellow) and IRR320 (gray), overlaid with their peaks arising from reduction. 2B both M+3 and M+6, showing intensity of M+6 peak.

Appendix G: Supplemental Information for Chapter 6

G.1 Sample collection for and preparation of C_{18} extracts

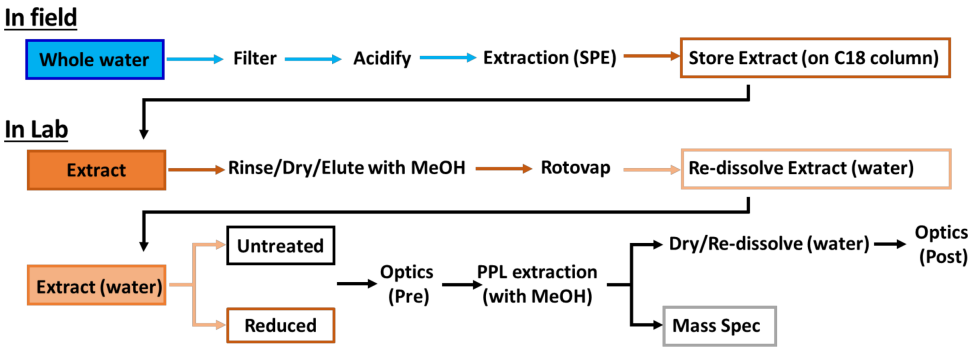


Figure G.1: Flow cart of the collection of natural water samples in the field and preparation for optical and MS experiments.

In field: Collection. Natural water samples were collected at selected depths by employing a surface pumping system (for surface samples) or conductivity, temperature, and depth (CTD) rosette equipped with Niskin bottles (for deep samples). Immediately after collection, samples were filtered through a 0.2 μm double layer HT Tuffryn hydrophilic polysulfone filter (NPO samples) or a 0.2 μm Gelman filter (EAO and DEL samples) and then acidified to pH 2 (with $\sim$100 mL of 2M HCl). To extract the DOM, the filtered acidified samples were pumped through C_{18} (United

Chemical Technologies) cartridges preconditioned with 100 mL MeOH and 50 mL pH 2 MQ water [10, 128] at a flow rate of 50 mL min^{-1} [44]. After the extraction, the cartridges were rinsed with MQ water (pH 2, ∼100 mL) to remove salts and then stored at -18°C in the dark until further processing.

In laboratory: Extraction The C_{18} cartridges were first rinsed with 0.1% formic acid (by volume in water) to remove any remaining HCl and then dried with a gentle nitrogen stream to remove water. DOM was extracted from all cartridges using methanol (50 mL). The first fraction (5mL) was kept separate from the second fraction (45 mL) to remove some of the aqueous residue; the second fraction was collected. The eluate was evaporated (by a rotary evaporator 35°C), re-dissolved in MQ water (∼2 mL), neutralized (NH_4OH) to pH ∼ 7, and then stored in the dark at -18°C until further processing and analysis.

North Pacific Ocean: NPO samples. Samples from the NPO (Station ALOHA) were collected onboard the R/V Kilo Moana during December 2014 at coordinates (22° 45' N, and 158° 00' W). A depth profile consisting of 11 depths from the surface down to approximately 4500 meters (includes several different water masses [129] Table G.1) were sampled using the method described above.

Table G.1: North Pacific Ocean Water Masses from Hernes et al. 2002 [129]

Water masses	Abbreviation	Depth (m)
North Pacific Sub-Tropical Water	NPSTW	0-200
Sub-tropical Mode Water	STMW	200-500
North Pacific Intermediate Water	NPIW	500-800
Antarctic Intermediate Water	AAIW	800-2000
North Pacific Deep Water	NPDW	2000-3000
Lower Circumpolar Water	LCPW (CDW)	3000-5000

Equatorial Atlantic Ocean: EAO samples. Samples from the Equatorial Atlantic Ocean were collected by Andrew et al. [10] onboard the R/V Endeavor in May-June of 2009 and were collected and processed as described above and as previously reported by Andrew et al. [10].

Delaware Transect: DEL samples. Samples from Delaware were collected by Boyle et al. [44] onboard the R/V Cape Henlopen in Aughust 2006 using the sample collection method described above.

G.2 North Pacific Ocean and Equatorial Atlantic Ocean Samples

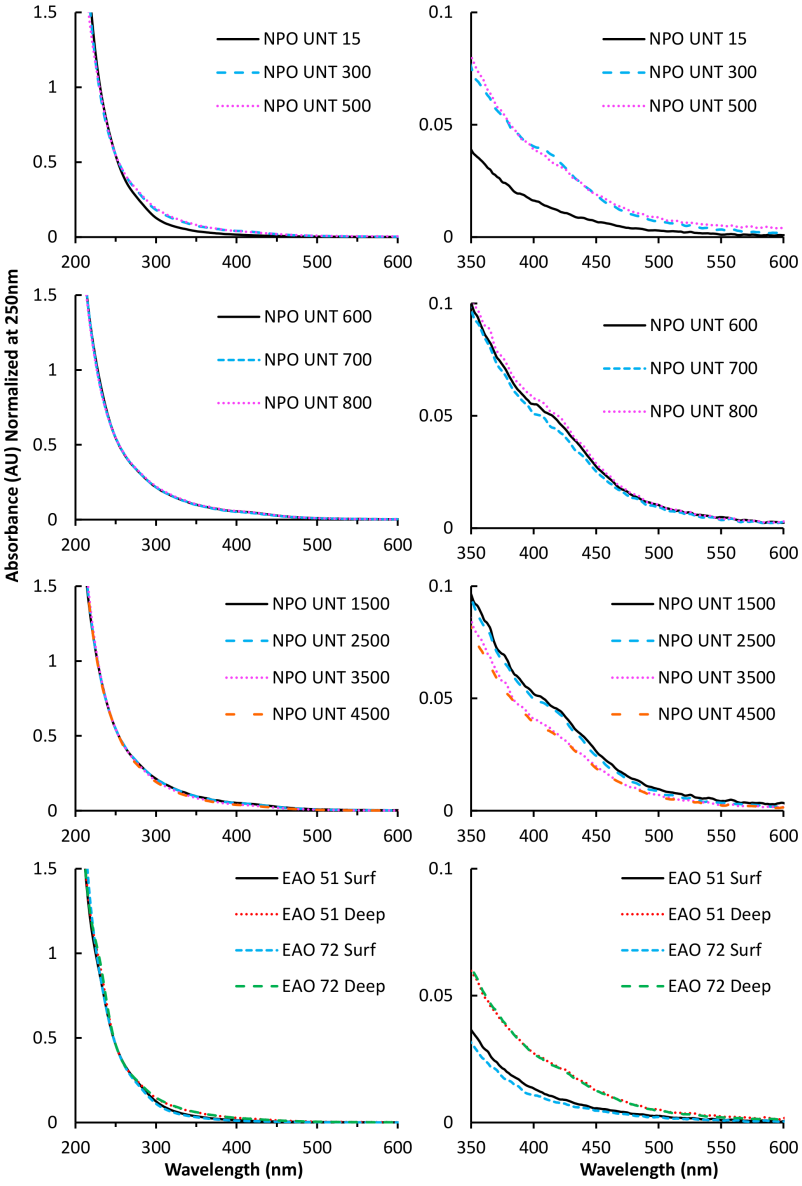


Figure G.2: Absorption of NPO and EAO samples Left) 200-600 nm; Right) zoomed in on 350-600 nm.

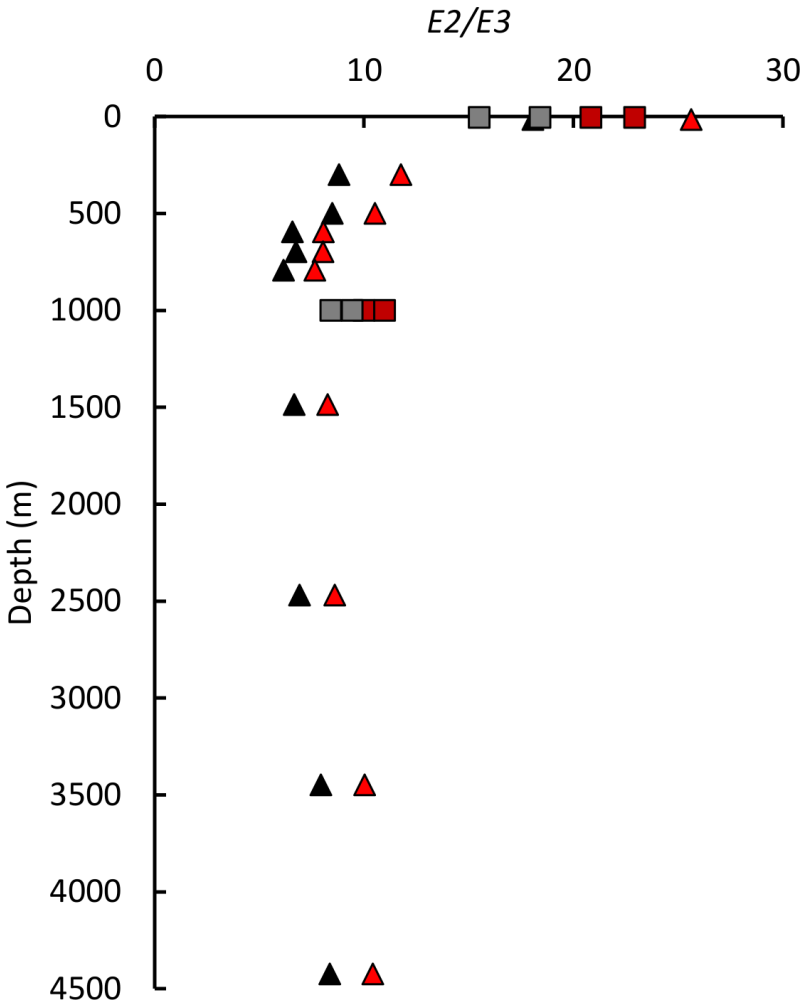


Figure G.3: $E2/E3$ ratio for untreated and $NaBD_4$ reduced for NPO and EAO samples. Untreated samples: NPO black triangles, EAO gray squares. Reduced: NPO red triangles, EAO dark red squares.

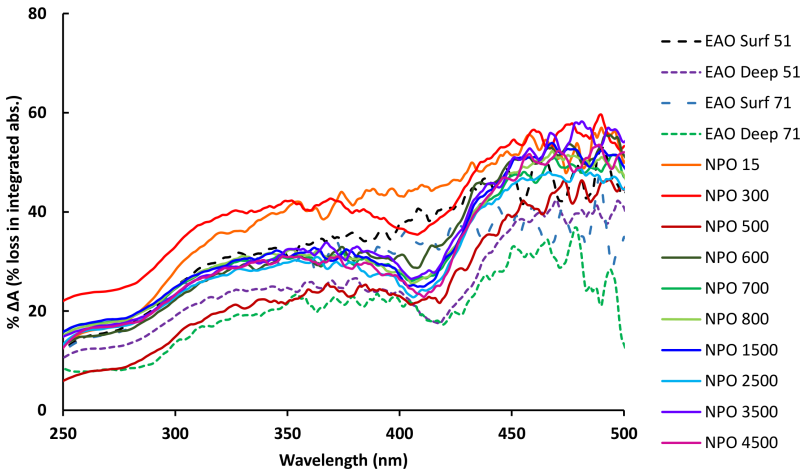


Figure G.4: Absorption loss after reduction by $NaBD_4$ for NPO and EAO samples

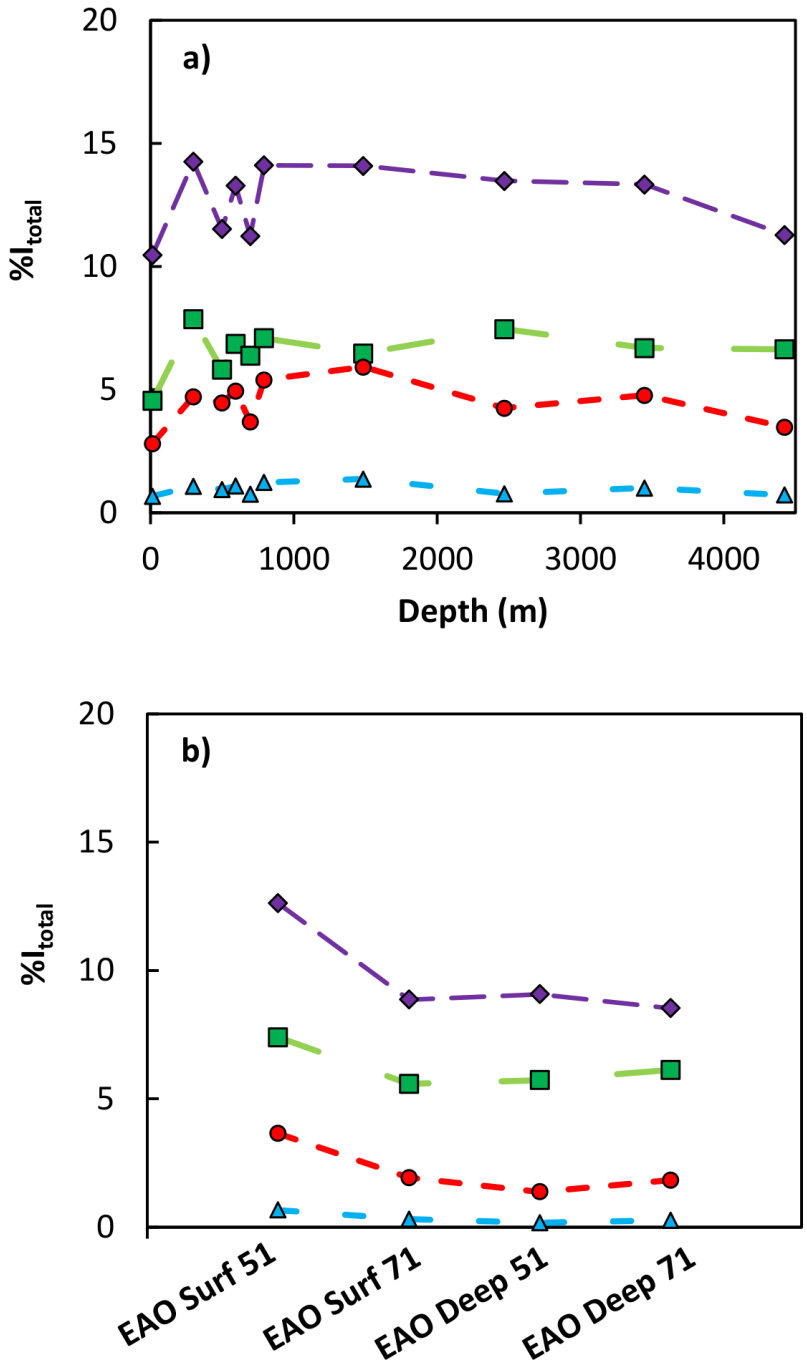


Figure G.5: $\%I_X$ extent of reduction vs a) NPO depth profile and b) EAO. $\%I_X$ (intensity-weighted percent of ions across 200-600 m/z). Total percent of reduced intensity (purple, $\%I_{total}$) allocated into groups: $\%I_{1O}$ (green) % of intensity of $M+3$ only peaks); $\%I_{1B}$ (red) both $M+3$ and $M+6$ found, showing % of intensity of $M+3$ peaks; $\%I_{2B}$ (blue) both $M+3$ and $M+6$ found, showing % of intensity of $M+6$ peaks; $\%I_{2O}$ not shown (% of intensity of $M+6$ only peaks account for less than 2% of the peak intensity for all samples).

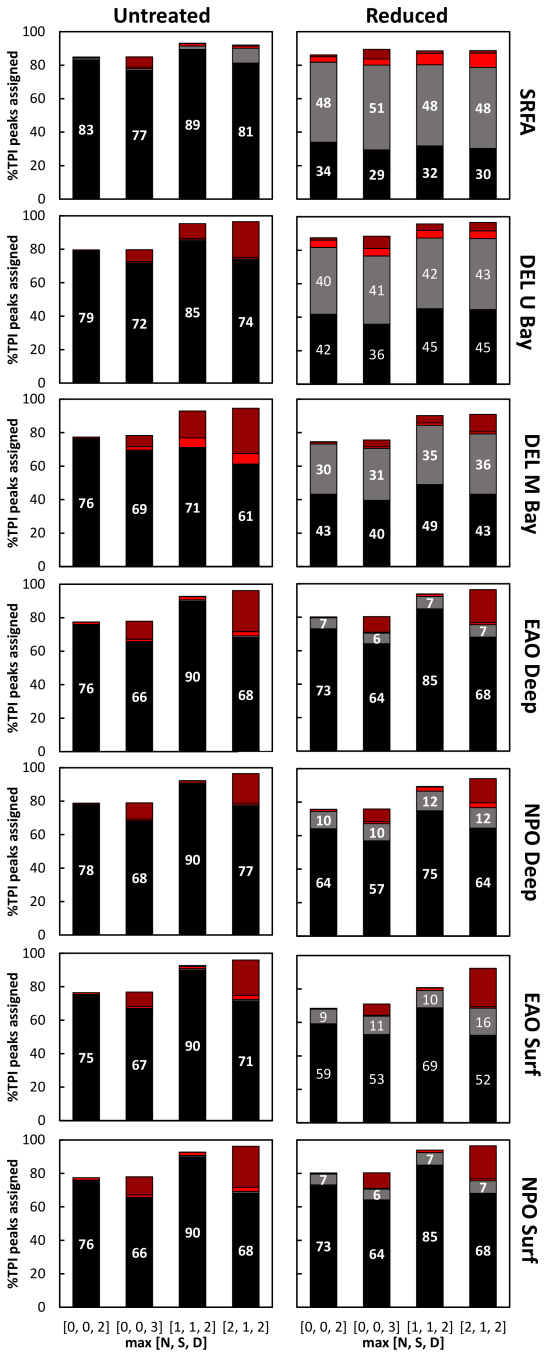


Figure G.6: Bar graph of %TPI for seven samples (untreated and reduced) assigned equivocal/unequivocal and D-containing/non-D-containing for various allowances of N, S, D (0,0,2; 0,0,3; 1,1,2; and 2,1,2, respectively) from 200-600 *m/z*. D=0 unequivalent (Black), D>0 unequivalent (gray), D=0 equivalent (red), D>0 equivalent (dark red).

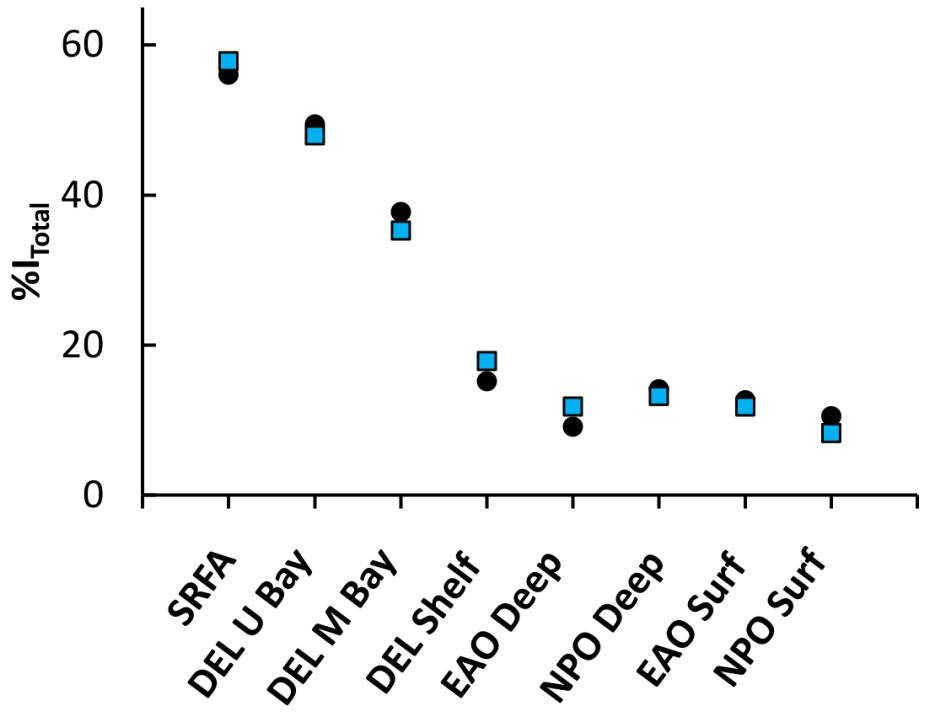


Figure G.7: Total percent of reduced intensity,$\%I_{total}$ (intensity-weighted percent of ions across 200-600 m/z, vs CHO (black circles) and $CHON_1S_1$ (blue squares) for different locals

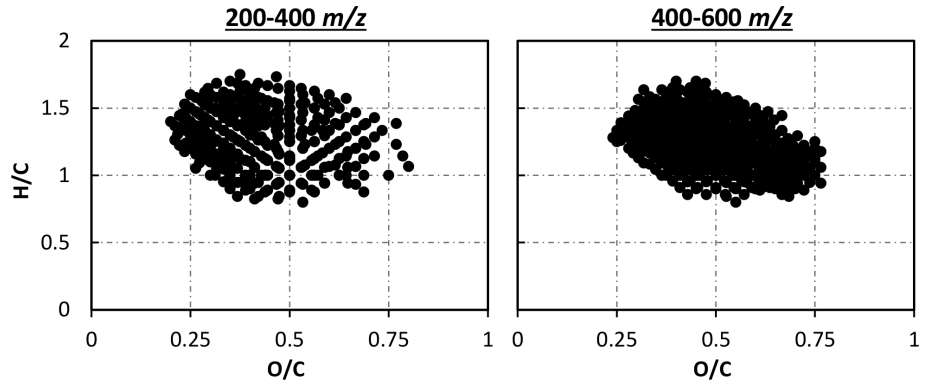


Figure G.8: Van Krevelen plot of the untreated molecular formulae all eight samples have in common split into two mass ranges.

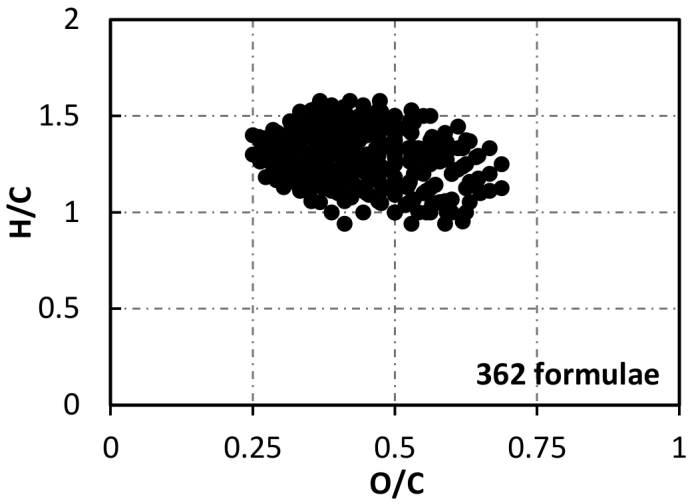


Figure G.9: Van Krevelen plot of the reduced molecular formulae that all eight samples have in common from 200-600 m/z.

Appendix H: MATLAB functions and descriptions

(Functions and scripts published in Bianca, M; Baluha, D; Gonsior, M; Schmitt-Kopplin, P; Del Vecchio, R; Blough, N.V. Contribution of ketone/ aldehyde-containing compounds to the composition and optical properties of Suwannee River fulvic acid revealed by ultrahigh resolution mass spectrometry and deuterium labeling. Anal. Bioanal. Chem. 2020, 412:1441–1451 [65].)

H.1 Preprocessing of raw mass spectral data (F1_list)

F1_list: output 'peaks_X'. Preprocessing of raw mass spectral data (raw_X). Blank (blk_major) subtracts and outputs list of singly charged ions (peaks_X).

```
function [rel_intpeaks, peaks, peaks_ct] = F1_list(mz_min, ...
    mz_max, raw, tol, blank)

%Pre-processes a raw mass spectral peak list:
%   * includes peaks in a specified m/z range
%   * removes doubly-charged peaks (if 'tol' is specified)
%   * blank subtracts peaks (if 'tol' and 'blank' is specified)
%   * displays # of peaks, total peak intensity, and average m/z

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%REQUIRED INPUT ARGUMENTS
%
% 'mz_min' and 'mz_max': Positive numbers specifying lower and ...
    upper limits
% of the the m/z range to be included in the final peak list
% (mz_min must be less than mz_max).
%
```

```
16 % 'raw': Two-column matrix containing the raw peak list of the mass
17 % spectrum of the sample to be analyzed, with m/z in the first ...
     column
18 % and raw intensities in the second column.
19 %
20 %OPTIONAL INPUT ARGUMENTS
21 %
22 % 'tol': Error limit (in ppm; positive number) used to find and ...
     excise
23 % peaks that are present in the blank spectrum and to search for ...
     and excise
24 % multiply-charged peaks.
25 %
26 % 'blank': Two-column matrix containing the raw peak list of the ...
     mass
27 % spectrum of a corresponding solvent blank, with m/z in the first
28 % column and raw intensities in the second column.
29 %
30 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
31 %OUTPUT ARGUMENTS
32 %
33 % 'peaks': Two-column matrix containing m/z (first column) and
34 % intensity data (second column) of the sample after the
35 % following processes are performed on the raw data:
36 %   (1) Excision of peaks outside the specified m/z range
37 %   (2) If 'tol' is given:
38 %          Excision of peaks which differ by 0.501678 within ...
     'tol' ppm
39 %          (i.e., the monoisotopic peak and the corresponding 13C
40 %          isotopologue). If more than one peak is found within ...
     'tol' ppm
41 %          of a theoretical isotopologue's m/z, the peak with ...
     the smallest
42 %          deviation from the theoretical value will be excised.
43 %
44 %   (3) If 'tol' and 'blank' are given:
45 %          Excision of peaks which have m/z values within 'tol' ...
     ppm of a
46 %          peak present in 'blank'.
47 %
48 % 'peaks_ct': matrix with the following rows, displayed when ...
     function ends
49 %
50 %  ((1) # peaks outside of m/z range excised
51 %  (2) # peaks with z=2 excised
52 %  (3) # peaks in blank excised
53 %  (4) number of peaks remaining
54 %  (5) total peak intensity (TPI)
55 %  (6) number-average m/z (AMN) = sum(mz*intensity)/sum(intensity)
56 %  (7) weight-average m/z (AMW) = ...
     sum(mz^2*intensity)/sum(mz*intensity)
57 %  (8) polydispersity = AMN / AMW
58 %  (9) m/z of base peak
59 % (10) raw intensity of base peak
```

```
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

peaks_ct = zeros(8,1);

%converts 'isnan' values to 0 (if needed)
raw(isnan(raw)) = 0;

for i = 1 : size(raw,1)
    %Excises peaks by setting both columns in 'raw' to [0 0]

    %Excises peaks with intensity equal to 0
    if raw(i,2) == 0
        raw(i,1:2)=zeros(1,2);
    end

    %Excises peaks outside the specified m/z range
    if raw(i,1) < mz_min || raw(i,1) > mz_max
        raw(i,1:2) = zeros(1,2);
     peaks_ct(1) = peaks_ct(1) + 1;
    end

    %Excises doubly-charged peaks if 'tol' is given
    if nargin > 3
        thr_mz = raw(i,1) + 0.501678; %theoretical m/z of ...
            isotopologue
        for k = 1 : size(thr_mz)
            %Finds peak with smallest deviation from thr_mass
            [err, z] = min( ...
                abs(1000000*(raw(:,1)-thr_mz(k))/thr_mz(k)));
            if err ≤ tol
                raw(i,1:2) = zeros(1,2); %base peak
                raw(z,1:2) = zeros(1,2); %isotopologue
          peaks_ct(2) = peaks_ct(2) + 2;
            end
        end
    end

    %Excises peaks shared in 'blank' (if given)
    if nargin > 4
        for k = 1 : size(blank,1)
            if abs( 1000000 *(raw(i,1) - blank(k,1)) / ...
                blank(k,1) ) ≤ tol
                raw(i,1:2) = zeros(1,2);
             peaks_ct(3) = peaks_ct(3) + 1;
            end
        end
    end
end

%stores non-excised values in 'PEAKS' (removes rows of zeros)
x = 0;
peaks = zeros(1,2);
for i = 1 : size(raw,1)
```

```
111     if raw(i,1) > 0
112         x = x + 1;
113         peaks(x,1) = raw(i,1);
114         peaks(x,2) = raw(i,2);
115     end
116 end
117 
118 %stores non-excised values in 'PEAKS' (removes rows of zeros)
119 %calcs rel intensity of peaks
120 
121 [peaks_ct(10), BP_index] = max(raw(:,2)); %m/z of base peak
122  peaks_ct(9) =raw(BP_index,1); %raw intensity of base peak
123 
124 rel_intpeaks=peaks;
125 x = 0;
126 rel_intpeaks = zeros(1,2);
127 for i = 1 : size(raw,1)
128     if raw(i,1) > 0
129         x = x + 1;
130         rel_intpeaks(x,1) = raw(i,1);
131         rel_intpeaks(x,2) = 100* raw(i,2) / peaks_ct(10);
132     end
133 end
134 
135 %number of peaks remaining
136 peaks_ct(4) = size(peaks,1);
137 
138 %total peak intensity (TPI)
139 peaks_ct(5) = sum(peaks(:,2));
140 
141 %number-average m/z (AMN) = sum(mz*intensity)/sum(intensity)
142 peaks_ct(6) = sum(peaks(:,1).*peaks(:,2)) / sum(peaks(:,2));
143 
144 %weight-average m/z (AMW) = sum(mz^2*intensity)/sum(mz*intensity)
145 
146 peaks_ct(7) = sum(peaks(:,1).*peaks(:,1).*peaks(:,2)) / ...
        sum(peaks(:,1).*peaks(:,2));
147 
148 %polydispersity = AMN / AMW
149 peaks_ct(8) = peaks_ct(6,1) / peaks_ct(7,1);
150 
151 disp(peaks_ct(1));
152 disp(peaks_ct(2));
153 disp(peaks_ct(3));
154 disp(peaks_ct(4));
155 disp(peaks_ct(5));
156 disp(peaks_ct(6));
157 disp(peaks_ct(7));
158 disp(peaks_ct(8));
159 
160  % 'peaks_ct': matrix with the following rows, displayed when ...
        function ends
161 %
162 %   (1) # peaks outside of m/z range excised
```

```
%   (2) # peaks with z=2 excised
%   (3) # peaks in blank excised
%   (4) number of peaks remaining
%   (5) total peak intensity (TPI)
%   (6) number-average m/z (AMN)
%   (7) weight-average m/z (AMW)
%   (8) polydispersity = AMN / AMW

end

%=====================================
% Daniel R. Baluha, PhD
% MATLAB version R2018
% June 18, 2019
%=====================================
```

H.2 Molecular formula assignment (F2_form)

F2_form: output 'MF_X'. Molecular formula calculated using "low mass moiety" approach from the output of F1_list (peaks_X).

```
function [MF, CT] = F2_form(peaks, tol, n_max, s_max, d_max)

%Calculates molecular formulae for m/z values in 'peaks'
%using a "low-mass moiety" algorithm as described by Green and ...
    Perdue in
%Anal. Chem. 2015, 87, 5079-5085 and
%Anal. Chem. 2015, 87, 5086-5094

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%REQUIRED INPUT ARGUMENTS
%
% 'MF_X': Matrix of peaks to which a molecular formula has been
% assigned with the following columns (see function 'F2_form' ...
    for more
% details):

%
% 'tol': Error threshold (in ppm) allowed for molecular formula ...
    assignment
%
%OPTIONAL INPUT ARGUMENTS
%
% 'n_max': maximum number of 14N atoms allowed. Default value = 0.
%
% 's_max': maximum number of 32S atoms allowed. Default value = 0.
```

```
23 %
24 % 'd_max': maximum number of deuterium (2H) atoms allowed.
25 % Default value = 0.
26 %
27 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
28 %OUTPUT ARGUMENT
29 %
30 % 'FORM': Matrix of peaks to which a molecular formula has been
31 % assigned with the following columns:
32 %
33 %         ....Number of atoms ...
34 % [mz int 12C 13C H   D   O   N   32S 34S FE  num O/C H/C]
35 %  1  2   3   4   5   6   7   8   9   10  11  12  13  14
36 %
37 %    *formula error (in ppm)
38 %
39 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
40 %EXPLANATION OF MOLECULAR FORMULA ASSIGNMENT
41 %
42 %PART 1. Each measured m/z is converted to its neutral mass (by ...
      addition of
43 %the mass of H+), and a base molecular formula of a hydrocarbon ...
      with the
44 %lowest possible H/C molar ratio is calculated for that nominal ...
      mass.
45 %This base formula is altered by adding or subtracting C4O-3 ...
      and/or CH4O-1
46 %subunits until the calculated mass of the formula is within ...
      'tol' ppm of
47 %the measured mass. Heteroatoms are included by varying the ...
      number of 32S
48 %and/or NH subunits to the base hydrocarbon. The following ...
      constraints must
49 %be satisfied:
50 %
51 %   2 ≤ #H ≤ 2(#C) + 2
52 %   O ≤ C
53 %   N  ≤ (user specified)
54 %   S ≤ (user specified)
55 %   O+N+S > 0
56 %
57 %The difference in ppm between the experimentally measured m/z (MM)
58 %and the theoretical m/z of each formula (CM) is calculated ...
      using the
59 %following equation:
60 %
61 %   FE = 1000000*(MM-CM)/CM
62 %
63 %   where CM = 12*C + 1.007825*H + 2.014102*D + 15.994915*O +
64 %   14.003074*N + 33.967867*S
65 %
66 %   If the absolute value of FE is less than or equal to the error
67 %   threshold specified by 'tol', the formula is assumed to be ...
      correct.
```

```
%
%PART 2. +13C-12C and +34S-32S isotopologue peaks are searched ... for each
%peak to which a molecular formula was assigned. If a peak whose ... measured
%mass is within 'tol' ppm of the calculated mass of the isotopologue
%peaks is found, the formula of this isotopologue is used.
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

tic

 if nargin < 5
    d_max = 0;
    if nargin < 4
        s_max = 0;
        if nargin < 3
            n_max = 0;
        end
    end
end

%places m/z and intensity data into matrix 'FORM1'
FORM1 = zeros( size(peaks,1),18 );
FORM1(:,1) = peaks(:,1);
FORM1(:,2) = peaks(:,2);

% [mz int 12C 13C 1H  2D  16O 14N 32S 34S FE  num O/C H/C]
%  1  2   3   4   5   6   7   8   9   10  11  12  13  14


%exact atomic masses of 12C, 1H, 2H, 16O, 14N, 32S,
EAM = [12; 13.0033548; 1.0078250; 2.0141018; 15.9949146; ... 14.0030740; 31.9720710; 33.9678669];

%[12; 13.0033548; 1.007825; 2.014102; 15.994915; 14.003074; ... 31.972071; 33.967867];

for x = 1 : size(FORM1,1)
    %convert to neutral mass
    NM = FORM1(x,1) + 1.0072764;

    %cycle through number of 13C 2H, 14N, 32S  containing moieties
    for d = 0: d_max %:-1: 0   %number of 1H+2H
        for s = 0: s_max %:-1: 0   %number of 1H+14N
            for n = 0 : n_max  %:-1: 0   %number of 32S

                %formula for m/z 'x'
                form = zeros(1,8);
                %mass of heteroatom-containing moieties
                sum_HA = d*3.0219268 + s*31.972071 + n*15.010899;

                %core CH mass
                MMc = NM - sum_HA;
```

```
118
119                 if mod(round(MMc),2) == 0
120                     i = 0;
121                     %initial (lowest H/C) hydrocarbon formula
122                     form(1) = floor(MMc/12); %highest C# possible
123                     form(2) = 0;
124                     form(3) = round(MMc - 12*form(1)) + n + d;
125                     form(4) = d;
126                     form(5) = 0;
127                     form(6) = n;
128                     form(7) = s;
129                     form(8) = 0;
130
131                     while i < 100
132                         %calculate current formula mass
133                         %CFM = form * EAM;
134                         %calculate formula error
135                         FE = 1000000*(NM - form*EAM)/(form*EAM);
136
137                         if abs(FE) ≤ tol
138                             %end 'while' loop, check formula
139                             i = 100;
140
141                         else
142                             % number of CH4-O replacements needed
143                             M = round((NM - form*EAM)/0.0363854);
144
145                             %formula error after M CH4-O ...
                                replacements
146     FE = 1000000*(NM-form*EAM-M*0.0363854)/((form*EAM));
147
148                             if abs(FE) ≤ tol
149                                 %perform M +CH4-O replacements
150                                 form(1) = form(1) + M;
151                                 form(3) = form(3) + 4*M;
152                                 form(5) = form(5) - M;
153                                 %end 'while' loop, check formula
154                                 i = 100;
155                             else
156                                 %perform +O3-C4 replacement
157                                 form(1) = form(1) - 4;
158                                 form(5) = form(5) + 3;
159                                 i = i + 1;
160                             end
161                         end
162
163                         FE = 1000000*(NM - form*EAM) / (form*EAM);
164
165                         %formula check: within error
166                         if abs(FE) ≤ tol
167
168                             %formula checks: C > 0, 2 ≤ H ≤ 2C + 2
169                             if form(1) > 0  && form(3) ≥ 2 && ...
                                form(3) ≤ 2*form(1)+2
```

```
170
171                                 %formula check: O+N+S > 0
172                                 if sum(form(5:8))>0
173
174                                     %formula check: 0 ≤ O ≤ C
175                                     if form(5) ≤ form(1) && ...
                                            form(5)≥0
176
177                                         FORM1(x,12) = ...
                                                FORM1(x,12) + 1;
178                                         FORM1(x,3:10) = form;
179                                         FORM1(x,11) = FE;
180
181                                         %13C-12C isotopologue
182                                         CMci = form*EAM + ...
                                                1.0033548 - 1.0072764;
183         [FEci, cz] = min(abs(1000000*(FORM1(:,1)-CMci)/CMci));
184                                         if FEci ≤ tol
185                                             FORM1(cz,3:10) = ...
                                                    FORM1(x,3:10);
186                                             FORM1(cz,3) = ...
                                                    FORM1(cz,3)-1;
187                                             FORM1(cz,4) = 1;
188                                             FORM1(cz,11) = FEci;
189                                             FORM1(cz,12) = ...
                                                    FORM1(cz,12) + 1;
190                                         end
191
192                                         %34S-32S isotopologue
193                                         if FORM1(x,9) > 0
194                                             CMsi = form*EAM + ...
                                                    1.995796 - ...
                                                    1.0072764;
195         [FEsi, sz] = min(abs(1000000*(FORM1(:,1)-CMsi)/CMsi));
196                                             if FEsi ≤ tol
197                                                 FORM1(sz,3:10) = ...
                                                        FORM1(x,3:10);
198                                                 FORM1(sz,9) = ...
                                                        FORM1(sz,9)-1;
199                                                 FORM1(sz,10) = 1;
200                                                 FORM1(sz,11) = FEsi;
201                                                 FORM1(sz,12) = ...
                                                        FORM1(sz,12) ...
                                                        + 1;
202                                             end
203                                         end
204                                     end
205                                 end
206                             end
207                         end
208                     end
209                 end
210             end
211          end
```

```
212     end
213 end
214
215  %Deletes rows with no formula
216 MF = zeros(1,14);
217 CT = zeros(3,2);
218
219 % row 1: all formula
220 % row 2: unequivocal
221 % row 3:D > 0
222 % col 1: number
223 % col 2: peak intensity
224
225  for i=1:size(FORM1,1)
226     if sum(FORM1(i,3:10)) > 0
227     CT(1,1) = CT(1,1) + 1;
228     CT(1,2) = CT(1,2) + FORM1(i, 2);
229         MF(CT(1,1),1:12) = FORM1(i, 1:12);
230     if FORM1(i,12) == 1
231         CT(2,1) = CT(2,1) + 1;
232         CT(2,2) = CT(2,2) + FORM1(i, 2);
233     end
234     if FORM1(i,6) > 0
235         CT(3,1) = CT(3,1) + 1;
236         CT(3,2) = CT(3,2) + FORM1(i, 2);
237     end
238     end
239 end
240
241 CT(:,2)=100*CT(:,2)/sum(peaks(:,2));
242
243 if size(MF,1)>1
244     disp(['total .....N = ',num2str(CT(1,1)),'; %I =' ...
            num2str(CT(1,2))]);
245     disp(['unequiv ...N = ',num2str(CT(2,1)),'; %I =' ...
            num2str(CT(2,2))]);
246     disp(['D > 0 .....N = ',num2str(CT(3,1)),'; %I =' ...
            num2str(CT(3,2))]);
247 end
248
249  %calculates H/C of each formula
250 MF(:,13)=MF(:,7)./(MF(:,3)+MF(:,4));
251
252 %calculates O/C of each formula
253 MF(:,14)=(MF(:,5)+MF(:,6))./(MF(:,3)+MF(:,4));
254
255 %calculates molar ratios, etc for each formula
256 for i = 1 : size(MF,1)
257     if sum(MF(i,3:10)) > 0
258         c = MF(i,3) + MF(i,4);
259         h = MF(i,5) + MF(i,6);
260         o = MF(i,7);
261         n = MF(i,8);
262         s = MF(i,9) + FORM1(i,10);
```

```
263
264 %calculates AI
265         MF(i,15) = (1+c-o-s-(0.5*h)) / (c-o-s-n); %AI
266         if MF(i,15) < 0 || isnan(MF(i,15)) || isinf(MF(i,15));
267             MF(i,15) = 0;
268         end
269
270 %calculates DBE and DBE-O
271         MF(i,16) = c + 1 - (h/2) + (n/2); %DBE
272         MF(i,17) = c + 1 - (h/2) + (n/2) - o; %DBE-O
273     end
274 end
275 end
276
277  %=====================================
278 % Daniel R. Baluha, PhD
279 % MATLAB version 2018
280 % Aug. 13, 2019
281 %=====================================
```

H.3 Search (F3_search)

F3_search: output 'MFR_X.' Uses molecular formula calculated for the untreated sample from F2_form (MF_UNT) and the peaks from the reduced sample (peaks_RED) to label what molecular formula have been reduced.

```
1  function [MFR, MFR_T, RED, RED_T] = F3_search(MF_X, peaks, tol, ...
       n_max, s_max, d_max)
2
3  %Identifies species in 'form' which comprise species with ...
       borodeuteride
4  %reducible moieties (i.e. ketones, aldehydes) by searching ...
       M+3.021927 &
5  %M+6.043854 m/z in 'peaks'
6  %
7  %Calculates molecular formulae for unassigned masses in 'peaks' ...
       by calling
8  %function 'F2_form'
9
10 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
11 % ABBREVIATIONS:
12 %    "M+3" = M + 3.021 926 8
13 %    "M+6" = M + 6.043 853 6
14 %
```

```
15 % where M refers to an m/z in 'form'
16 %
17 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
18 %INPUT ARGUMENTS
19 %
20 % 'MF_X': Matrix of peaks to which a molecular formula has been
21 % assigned with the following columns (see function 'F2_form' ...
      for more
22 % details):
23 %
24 %          ...Number of atoms ...
25 % [mz int 12C 13C H D O N 32S 34S FE O/C H/C N/C AI  DBE DBE-O]
26 %  1  2   3   4   5 6 7 8   9 10  11 12  13  14  15  16  17
27 %
28 % 'peaks': Matrix containing at least two columns, with the first
29 % column containing m/z values and the second column containing ...
      intensities
30 % (should be a borodeuteride-reduced mass spectral peak list).
31 %
32 %
33 % 'tol': Error threshold (in ppm) allowed for molecular formula ...
      assignment
34 % and M+3/6 searches
35 %
36 %OPTIONAL INPUT ARGUMENTS
37 %
38 % 'n_max': maximum number of 14N atoms allowed. Default value = 0.
39 %
40 % 's_max': maximum number of 32S atoms allowed. Default value = 0.
41 %
42 % 'd_max': maximum number of deuterium (2H) atoms allowed.
43 %  Default value = 2.
44 %
45 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
46 %OUTPUT ARGUMENTS
47 %
48 % 'MFR': Same format as 'MF_X' with column 18 data added, which ...
      identifies
49 % the peak composition based on presence/absence of M+3 or M+6 ...
      masses in
50 % the searched mass spectral peak list. Values and meanings are ...
      as follows:
51 %     0 : non-reduced species only (default)
52 %     1 : singly-reduced species only
53 %     2 : doubly-reduced species only
54 %     3 : singly and doubly reduced species
55 % theoretical M+3 and M+6 masses are determined from the calculated
56 % (theoretical) mass of the given molecular formula. An M+3 or ...
      M+6 mass
57 % marker is considered to be present if there is a peak with m/z ...
      within
58 % 'tol' ppm of the theoretical M+3 or M+6 mass.
59 %
60 % 'MFR_T': Table with rows and columns defined as follows:
```

```
%
%    Columns: Peaks comprising species which are ___ reduced:
%            (1) non
%            (2) singly-reduced only
%            (3) doubly-reduced only
%            (4) singly- & doubly-reduced only
%
%
% 'RED': Same format as 'MF_RED' from function 'F2_form'
%
% 'RED_T': Same format as 'FORM_T' from function 'F2_form'
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
tic

 if nargin < 6
    d_max = 2;
    if nargin < 5
        s_max = 0;
        if nargin < 4
            n_max = 0;
        end
    end
end

MFR = MF_X;
MFR(:,18) = zeros(size(MF_X,1),1);
RED_temp(1:size(peaks,1),1:2) = peaks(:,1:2);

if nargout > 2
    RED_temp(1:size(peaks,1),3:17) = zeros(size(peaks,1),15);
end

Masses = [12; 13.0033548; 1.0078250; 2.0141018; 15.9949146; ...
    14.0030740; 31.9720710; 33.9678669];

for i = 1:size(MFR,1);
    %calculated M minus the mass of H+
    M = MFR(i,3:10)*Masses - 1.0072764;
    %M = LAB(i,1);
    M3 = M + 3.0219268; %theoretical M+3 mass
    M6 = M + 6.0438536; %theoretical M+6 mass
    if d_max ≥ 1
        %Search for m/z in 'RED_temp' closest to theoretical M+3
        [M3_diff, M3_in] = min( abs( M3 - RED_temp(:,1) ) );
        %if M+3 found & formula contains at least 1 O or N AND ...
            at least 1 DBE...
        if tol ≥ 1000000 * M3_diff / M3 && sum(MFR(i,7:8)) > 0 ...
            && MFR(i,16) > 0
            %increase ketone/aldehyde ID by 1
            MFR(i,18) = MFR(i,18) + 1;
            %if formulae are to be calculated for 'peaks'...
            if nargout > 2
                %mass shift-based molecular formula stored in ...
```

```
                    'RED_temp'
112                 RED_temp(M3_in,3:17) = MFR(i,3:17);
113                 RED_temp(M3_in,5) = RED_temp(M3_in,5) + 1;
114                 RED_temp(M3_in,6) = 1;
115                 calc = (RED_temp(M3_in,3:10)*Masses)-1.0072764;
116                 meas = RED_temp(M3_in,1);
117                 RED_temp(M3_in,11) = 1000000 * (meas - calc)/calc;
118                 RED_temp(M3_in,13) = sum(RED_temp(M3_in,5:6)) / ...
                        sum(RED_temp(M3_in,3:4));
119                 AI_d = 0.5 + sum(RED_temp(M3_in,3:4)) - ...
                        RED_temp(M3_in,7) - ...
                        sum(RED_temp(M3_in,9:10)) - RED_temp(M3_in,5)/2;
120                 AI_c = sum(RED_temp(M3_in,3:4)) - ...
                        RED_temp(M3_in,7) - ...
                        sum(RED_temp(M3_in,9:10)) - RED_temp(M3_in,8);
121                 RED_temp(M3_in,15) = AI_d / AI_c;
122                 RED_temp(M3_in,16) = RED_temp(M3_in,16) - 1;
123                 RED_temp(M3_in,17) = RED_temp(M3_in,17) - 1;
124             end
125         end
126     end
127
128     if d_max ≥ 2
129         %Search for m/z in 'RED_temp' closest to theoretical M+6
130         [M6_diff, M6_in] = min( abs( M6 - RED_temp(:,1) ) );
131         %if M+6 found & formula contains at least 2 O or N AND ...
                at least 2 DBE...
132         if tol ≥ 1000000 * M6_diff / M6 && sum(MFR(i,7:8)) > 1 ...
                && MFR(i,16) > 1
133             %increase ketone/aldehyde ID by 2
134             MFR(i,18) = MFR(i,18) + 2;
135             %if formulae are to be calculated for 'peaks'...
136             if nargout > 2
137                 %mass shift-based molecular formula stored in ...
                        'RED_temp'
138                 RED_temp(M6_in,3:17) = MFR(i,3:17);
139                 RED_temp(M6_in,5) = RED_temp(M6_in,5) + 2;
140                 RED_temp(M6_in,6) = 2;
141                 calc = (RED_temp(M6_in,3:10)*Masses)-1.0072764;
142                 meas = RED_temp(M6_in,1);
143                 RED_temp(M6_in,11) = 1000000 * (meas - calc)/calc;
144                 RED_temp(M6_in,13) = sum(RED_temp(M6_in,5:6)) / ...
                        sum(RED_temp(M6_in,3:4));
145                 AI_d = 1 + sum(RED_temp(M6_in,3:4)) - ...
                        RED_temp(M6_in,7) - ...
                        sum(RED_temp(M6_in,9:10)) - ...
                        RED_temp(M6_in,5)/2 - 1;
146                 AI_c = sum(RED_temp(M6_in,3:4)) - ...
                        RED_temp(M6_in,7) - ...
                        sum(RED_temp(M6_in,9:10)) - RED_temp(M6_in,8);
147                 RED_temp(M6_in,15) = AI_d / AI_c;
148                 RED_temp(M6_in,16) = RED_temp(M6_in,16) - 2;
149                 RED_temp(M6_in,17) = RED_temp(M6_in,17) - 2;
150             end
```

```
        end
    end
end

%Creates 'LAB_T'
num  = zeros(4,1);
for i = 1:size(MFR,1);
    if MFR(i,18) == 0
        num(1) = num(1) + 1;
    elseif MFR(i,18) == 1
        num(2) = num(2) + 1;
    elseif MFR(i,18) == 2
        num(3) = num(3) + 1;
    elseif MFR(i,18) == 3
        num(4) = num(4) + 1;
    end
end
perc = 100 * num / sum(num);
Categories = {'non      '
              'singly   '
              'doubly   '
              'both     '};
MFR_T = table(num, perc, 'RowNames',Categories);
disp(MFR_T);


%Calculates molecular formulae for remaining peaks in 'RED_temp'
if nargout > 2
    [RED, RED_T] = F2_form_04_01(RED_temp, tol, n_max, s_max, 0);
end

end

%=====================================
% Daniel R. Baluha
% University of Maryland, College Park
% June 2015
% MATLAB version R2013b
% OS X version 10.9.5
%=====================================
```

H.4 Find Reduced peaks (F4_Rint)

F4_Rint: output 'Rint'. Finds reduced peaks and respective intensities.

```
function [ALIGNED_M3_RED_sorted,ALIGNED_M6_RED_sorted] = ...
    F4_Rint(MFR,RIpeaksRED)
```

```
2
3  %Identifies M+6 and M+6 peaks(and respective intensity)from MFR
4  %
5  %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
6  % ABBREVIATIONS:
7  %   "M+3" = M + 3.021 926 8
8  %   "M+6" = M + 6.043 853 6
9  % where M refers to an m/z in 'MFR'
10 %
11 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
12 %INPUT ARGUMENTS
13 %
14 % 'RIpeaksRED': peak list of the reduced sample from F1_list (2 ...
       columns)
15 % column 1- m/z values; column 2- intensities (of reduced sample)
16 %
17 % 'MFR_RED': 18 columns. uses columns 1,2, and 18.
18 %column 1- m/z , column 2 intensity
19 %column 18 labeled as having borodeuteride reducible moieties
20 %
21 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
22 %OUTPUT ARGUMENTS
23 %
24 % 'ALIGNED_M3_RED_sorted': Three-columns
25 %column 1- M (m/z of the species labeled as reduced)
26 %column 2- M+3(m/z of the reduced species)
27 %column 3-intensity of the reduced peak (int of M+3 peak)
28 %
29 %'ALIGNED_M6_RED_sorted': Three-columns
30 %column 1- M (m/z of the species labeled as reduced)
31 %column 2- M+6(m/z of the reduced species)
32 %column 3-intensity of the reduced peak (int of M+6 peak)
33 %
34 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
35
36 RLAB=MFR;
37
38 col1=getcolumn(RLAB(:,1:end),1);
39 col2=getcolumn(RLAB(:,2:end),1);
40 col18=getcolumn(RLAB(:,18:end),1);
41
42 Red_Lab=[col1,col2,col18];
43
44 clear col1 col2 col18
45
46 R2_PEAKS=RIpeaksRED;
47
48 PEAKS1 = Red_Lab;
49 PEAKS2 = Red_Lab;
50 PEAKS3 = Red_Lab;
51
52 for i = 1 : size(PEAKS1,1)
53             if PEAKS1(i,3) == 1;
54                    PEAKS1(i,4)=3.0219268;
```

```
            end
end

for i = 1 : size(PEAKS2,1)
               if PEAKS2(i,3) == 2;
                        PEAKS2(i,4)=6.0438536;
               else
                    PEAKS2(i,4)=0;
               end


end

for i = 1 : size(PEAKS3,1)
                if PEAKS3(i,3) == 3;
                        PEAKS3(i,4)=3.0219268;
                        if PEAKS3(i,3) == 3;
                            PEAKS3(i,5)=6.0438536;
                        end
                end
end

for i = 1 : size(PEAKS1,1)
    if PEAKS1(i,4) == 3.0219268;
        PEAKS1(i,6)=PEAKS1(i,1)+3.0219268;
    end
end

for i = 1 : size(PEAKS2,1)
    if PEAKS2(i,4) == 6.0438536;
        PEAKS2(i,6)=6.0438536+PEAKS2(i,1);
    else
        PEAKS2(i,6)=0
    end
end

for i = 1 : size(PEAKS3,1)
    if PEAKS3(i,4) == 3.0219268;
        PEAKS3(i,6)=PEAKS1(i,1)+3.0219268;
    if PEAKS3(i,5) == 6.0438536;
        PEAKS3(i,7)=6.0438536+PEAKS3(i,1);
    end
    end
end

PEAKS123=([PEAKS1(:,1),PEAKS1(:,6),PEAKS2(:,6),PEAKS3(:,6),...
PEAKS3(:,7)]);

PEAKS_Temp=PEAKS123;

for i = 1 : size(PEAKS_Temp,1)
    if PEAKS_Temp(i,2) ==0;
       PEAKS_Temp(i,2)=PEAKS_Temp(i,4);
    end
```

```
end

PEAKS_Temp(:,4)=0;


for i = 1 : size(PEAKS_Temp,1)
    if PEAKS_Temp(i,3) ==0;
       PEAKS_Temp(i,3)=PEAKS_Temp(i,5);
    end
end

PEAKS_Temp(:,5)=0;

PEAKS_M3andM6_mz=([PEAKS_Temp(:,1),PEAKS_Temp(:,2), ...
   PEAKS_Temp(:,3)]);

P=PEAKS_M3andM6_mz;

for i = 1 : size(P,1)
    if P(i,2)==0
       P(i,2)=i+1000;
    end
end

for i = 1 : size(P,1)
    if P(i,3)==0
           P(i,3)=i+1000;
    end
end

PEAKS_sepOrigM=P(:,1);
PEAKS_sepM3=P(:,2);
PEAKS_sepM6=P(:,3);

PeaksOM=([PEAKS_sepOrigM(:),PEAKS_sepOrigM(:)]);
PeaksM3=([PEAKS_sepM3(:),PEAKS_sepOrigM(:)]);
PeaksM6=([PEAKS_sepM6(:),PEAKS_sepOrigM(:)]);

[ALIGNED_OM_RED] = ms_compare_mod(0.4,PeaksOM,R2_PEAKS);
[ALIGNED_M3_RED] = ms_compare_mod(0.4,PeaksM3,R2_PEAKS);
[ALIGNED_M6_RED] = ms_compare_mod(0.4,PeaksM6,R2_PEAKS);

ALIGNED_OM_RED(ALIGNED_OM_RED(:,2)<0.01,:)=[];
ALIGNED_M3_RED(ALIGNED_M3_RED(:,2)<0.01,:)=[];
ALIGNED_M6_RED(ALIGNED_M6_RED(:,2)<0.01,:)=[];

[¬,idx]=sort(ALIGNED_M3_RED(:,2));
ALIGNED_M3_RED_sorted=ALIGNED_M3_RED(idx,:);

[¬,idx]=sort(ALIGNED_M6_RED(:,2));
ALIGNED_M6_RED_sorted=ALIGNED_M6_RED(idx,:);

for i = 1 : size(ALIGNED_M3_RED_sorted,1)
    if ALIGNED_M3_RED_sorted(i,1)>1000
```

```
            ALIGNED_M3_RED_sorted(i,1)=0;
    end
end

for i = 1 : size(ALIGNED_M6_RED_sorted,1)
    if ALIGNED_M6_RED_sorted(i,1)>1000
            ALIGNED_M6_RED_sorted(i,1)=0;
    end
end

%Note lines 143-145 use function ms_compare
%=====================================
% Daniel R. Baluha
% University of Maryland, College Park
% June 2015
% MATLAB version R2013b
% OS X version 10.9.5
%=====================================

%=====================================
% Marla Bianca
% University of Maryland, College Park
% 2019
% MATLAB version R2015b
%=====================================
```

H.5 Compare Script

ms_compare_mod used to sort and align data in function F4_Rint.

```
function [ALIGNED] = ms_compare_mod( tol, varargin )

%Aligns mz
TEMP = 0;
start = 1;
for i = 1 : nargin-1
    out = varargin{i};
    out_rows = size(out,1);
    out_cols = size(out,2);
    stop = start + out_rows - 1;
    if out_cols == 2 %col1 m/z col2 int
        alignd_cols = 1; %m/z
        TEMP(start:stop,1) = out(:,1);
        TEMP(start:stop,i+1) = out(:,2);
        start = stop+1;
    end
end
```

```
18 TEMP = sortrows(TEMP, 1);
19 ALIGNED = zeros(1, size(TEMP,2));
20 n_aligned = 1;
21 curr_strt = 1;
22 test_row = 1;
23 while test_row < size(TEMP,1);
24     test_row = test_row + 1;
25     if alignd_cols == 1;
26         diff = 1000000 * ( TEMP(test_row,1) - TEMP(curr_strt,1) ...
               ) / TEMP(curr_strt,1);
27         if diff > tol
28             %adds aligned data to 'aligned'
29             ALIGNED (n_aligned, 1) = mean( TEMP ...
                   (curr_strt:test_row-1, 1) );
30             for spec = 1 : nargin-1
31                 ALIGNED (n_aligned, spec+1 ) = sum ( TEMP ...
                       (curr_strt:test_row-1, spec+1) );
32             end
33             %start new loop
34             n_aligned = n_aligned + 1;
35             curr_strt = test_row;
36             %in_rnge = 1;
37         end
38     end
39 end
40
41 %Note ms_compare_mod is modified from ms_compare by
42 %=====================================
43 % Daniel R. Baluha
44 % University of Maryland, College Park
45 % June 2015
46 % MATLAB version R2013b
47 % OS X version 10.9.5
48 %=====================================
```

Appendix I: MATLAB scripts: Fluoroscence

I.1 Fluorescence scripts for data collected on FluoroMax-4 luminescence spectrometer and converting intensity into QS units

(Scripts published in Bianca, M; Baluha, D; Gonsior, M; Schmitt-Kopplin, P; Del Vecchio, R; Blough, N.V. Contribution of ketone/ aldehyde-containing compounds to the composition and optical properties of Suwannee River fulvic acid revealed by ultrahigh resolution mass spectrometry and deuterium labeling. Anal. Bioanal. Chem. 2020, 412:1441–1451 [65].)

Fluorescence scripts for data collected on FluoroMax-4: blank substraction, smoothing by 5pt moving average, removal of scattering peaks and instrument masking.

```
%FLUORO MAX-4 script
%Uses raw .dat file from fluoro instrumnet
%em: 290-700nm :ex 240-600nm (every 10 nm)

%read in .dat files
delimiterIn='\t'; %gets into column form
headerlinesIn=2; %ignores the first 2 lines

name1=input('MQ: ');
M1=importdata(name1,delimiterIn,headerlinesIn);
MQ=M1.data;

name2=input('sample name: ');
```

```
F1=importdata(name2,delimiterIn,headerlinesIn);
F=F1.data;

clear name1 name2 delimiterIn headerlinesIn M1

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%
%MQ subtraction

sub=F-MQ;
F_MQ_sub=sub(:,2:end);

em=F(:,1); %gets emission wavelengths

clear F sub

figure(1); hold on
subplot(2,1,1);
plot(em, F_MQ_sub);
title('MQ sub');
hold off

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%
%smoothing data by 5 point moving average

Smooth_F_temp=F_MQ_sub;

Smooth_F=smooth(Smooth_F_temp(:),5);

F_smooth=reshape(Smooth_F,411,37);

figure(2); hold on
subplot(2,1,1);
plot(em,F_smooth );
title('smoothed');
hold off
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%
% Removes scatter peaks

C=F_smooth;

[rows,columns]=size(C);

for i=size(rows);
    for j=1:10;
        C(154+i*20:i*20+188,j)=0;
        i=i+1;
   end
end

D=C;

```

```
for i=size(rows)
    for j=1:7;
        D(200+i*21:i*21+235,j)=0;
        i=i+1
   end
end

E=D;

for i=size(rows)
    for j=4:11;
        E(E < 0) = 0;
        E(1+i*10:i*12+15,j)=0;
        i=i+1;
   end
end

for i =size(rows);
    for j=1:11;
F=E(:,j);
y=F;
x = 1:size(F);
%Interpolate zeros:
idx = y≠0;
yn(i,:)=interp1(x(idx),y(idx),x,'pchip');
i=i+1;
    end
end

yn_01=yn';

F_01=E(:,12:end);

F_02=[yn_01 F_01];


figure(3); hold on
subplot(2,1,1);
plot(em,F_02);
title('scatter removed');
hold off
%%
% masking from instument removed

A=F_02;

[rows,columns]=size(A);

for i=size(rows)
    for j=6:columns;
        A(1:i*10,j)=0;
        i=i+1;
   end
end
```

```
122
123 A(A==0) = NaN;
124
125 figure(4); hold on
126 subplot(2,1,1);
127 plot(em,A);
128 title('masking removed');
129 hold off
130 %%
131
132 fluoro_final=num2cell(A);
133 fluoro_final(cellfun(@isnan,fluoro_final))={[]};
134 Table=cell2table(fluoro_final);
135 writetable(Table,'fluoro_final.dat');
136
137 fluoro_final_zero=num2cell(A);
138 fluoro_final_zero(cellfun(@isnan,fluoro_final_zero))={[0]};
139 Table=cell2table(fluoro_final_zero);
140 writetable(Table,'fluoro_final_zero.dat');
141
142 %=====================================
143 % Marla Bianca
144 % University of Maryland, College Park
145 % 2019
146 % MATLAB version R2015b
147 %=====================================
```

Fluorescence script: converting intensity into QS units

```
1  %converts the fl intensity into QS units
2  %%
3  %to read in .dat files
4  delimiterIn='\t'; %gets into column form
5  headerlinesIn=2; %ignores the first 2 lines
6
7  name1=input('QS file name: ');
8  QS1=importdata(name1,delimiterIn,headerlinesIn);
9  QS=QS1.data;
10
11 name2=input('H2SO4 file name: ');
12 HS=importdata(name2,delimiterIn,headerlinesIn);
13 H2SO4=HS.data;
14
15 clear QS1 HS name1 name2 headerlinesIn delimiterIn
16 %%
17 QS_02=(QS(:,2));
18 QS_03=(QS_02(1:321,1));
19 clear QS_02 QS
20
21 H2SO4_02=(H2SO4(:,2));
22 H2SO4_03=(H2SO4_02(1:321,1));
```

```
clear H2SO4_02 H2SO4

QSsubH2SO4=QS_03-H2SO4_03;
QS_int=sum(QSsubH2SO4);

figure(10); hold on
subplot (2,2,1)
plot(forqs,QS_03)
title('QS')
subplot (2,2,2)
plot(forqs,H2SO4_03)
title('H2SO4')
subplot (2,2,3)
plot (forqs,QSsubH2SO4)
title('QS sub H2SO4')
hold off
%%
% A is the MQ subtracted and scattering removed fluor

Fluoro_in_QS_units=(A./QS_int/10); %for 10 ppb
Fluoro_in_QS_units(Fluoro_in_QS_units==0) = NaN;

%%
Fluoro_in_QS=num2cell(Fluoro_in_QS_units);
Fluoro_in_QS(cellfun(@isnan,Fluoro_in_QS))={[]};
Table=cell2table(Fluoro_in_QS);
writetable(Table,'srfa_titration_Fluoro_in_QS_units.dat');

%%
figure(11); hold on
subplot (2,1,1)
plot (em,A)
title('origional')

subplot (2,1,2)
plot(em,Fluoro_in_QS_units)
title('Fluoro in QS_units')
hold off
%=======================================
% Marla Bianca
% University of Maryland, College Park
% 2019
% MATLAB version R2015b
%=======================================
```

Bibliography

[1] J Swietlik, A D, U Raczyk-stanis, and J Nawrocki. Reactivity of natural organic matter fractions with chlorine dioxide and ozone. *Water Research*, 38:547–558, 2004.

[2] Patricia A Maurice, Michael J Pullin, Stephen E Cabaniss, Qunhui Zhou, Ksenija Namjesnik-dejanovic, and George R Aiken. A comparison of surface water natural organic matter in raw filtered water samples , XAD , and reverse osmosis isolates. *Water Research*, 36:2357–2371, 2002.

[3] S E Cabaniss and M S Shuman. Copper binding by dissolved organic matter : I . Suwannee River fulvic acid equilibria. *Geochimica et Cosmochimica Acta*, 52:185–193, 1988.

[4] Emma L Sharp, Simon A Parsons, and Bruce Jefferson. The impact of seasonal variations in DOC arising from a moorland peat catchment on coagulation with iron and aluminium salts. *Environmental Pollution*, 140:436–443, 2006.

[5] Yu Ping Chin, Penney L. Miller, Lingke Zeng, Kaelin Cawley, and Linda K. Weavers. Photosensitized degradation of bisphenol A by dissolved organic matter. *Environmental Science and Technology*, 38(22):5888–5894, 2004.

[6] Kelli S. Golanoski, Shuo Fang, Rossana Del Vecchio, and Neil V. Blough. Investigating the mechanism of phenol photooxidation by humic substances. *Environmental Science and Technology*, 46(7):3912–3920, 2012.

[7] Daniel R. Baluha. *A Deuterium Labeling Method for the Characterization of (Chromophoric) Dissolved Organic Matter Using Ultrahigh Resolution Electrospray Ionization Mass Spectrometry*. Ph.d. dissertation, University of Maryland College Park, 2015.

[8] Todd Pagano, Morgan Bida, and Jonathan E Kenny. Trends in Levels of Allochthonous Dissolved Organic Carbon in Natural Water: A Review of Potential Mechanisms under a Changing Climate. *Water*, 6(10):2862–2897, 2014.

[9] P.J. LeB Williams, M. Sondergaard, and D. Evans. Chapter 1 Dissloved organic matter (DOM): What is it and why study it? In David N Thomas and Morten Sondergaard, editors, *Dissolved Organic Matter (DOM) in Aquatic Ecosystems: A Study of European Catchments and Coastal Waters*, chapter Chapter 1, page 9. The Domaine Project, 2004.

[10] Andrea a. Andrew, Rossana Del Vecchio, Ajit Subramaniam, and Neil V Blough. Chromophoric dissolved organic matter (CDOM) in the Equatorial Atlantic Ocean: Optical properties and their relation to CDOM structure and source. *Marine Chemistry*, 148:33–43, 2013.

[11] Peter J Hernes and Ronald Benner. Photochemical and microbial degradation of dissolved lignin phenols : Implications for the fate of terrigenous dissolved organic matter in marine environments. *JOURNAL OF GEOPHYSICAL RESEARCH*, 108, 2003.

[12] Alexandra C Stenson, Alan G Marshall, and William T Cooper. Exact Masses and Chemical Formulas of Individual Suwannee River Fulvic Acids from Ultrahigh Resolution Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectra Molecular formulas have been assigned for 4626 indimass measurements fr. *Analytical Chemistry*, 75(6):1275–1284, 2003.

[13] Aron Stubbins, Robert G. M. Spencer, Hongmei Chen, Patrick G. Hatcher, Kenneth Mopper, Peter J. Hernes, Vincent L. Mwamba, Arthur M. Mangangu, Jose N. Wabakanghanzi, and Johan Six. Illuminated darkness: Molecular signatures of Congo River dissolved organic matter and its photochemical alteration as revealed by ultrahigh precision mass spectrometry. *Limnology and Oceanography*, 55(4):1467–1477, 2010.

[14] Nagamitsu Maie, Oliva Pisani, and Rudolf Jaffe. Mangrove tannins in aquatic ecosystems : Their fate and possible influence on dissolved organic carbon and nitrogen cycling Mangrove tannins in aquatic ecosystems : Their fate and possible influence on dissolved organic carbon and nitrogen cycling. *Limnology and Oceanography*, 53:160–171, 2008.

[15] Rachel L. Sleighter and Patrick G. Hatcher. Molecular characterization of dissolved organic matter (DOM) along a river to ocean transect of the lower Chesapeake Bay by ultrahigh resolution electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *Marine Chemistry*, 110(3-4):140–152, 2008.

[16] Mary Ann Moran, Wade M. Sheldon, and Richard G. Zepp. Carbon loss and optical property changes during long-term photochemical and biological degradation of estuarine dissolved organic matter. *Limnology and Oceanography*, 45(6):1254–1264, 2000.

[17] D J Repeta, Tracy M Quan, L I Aluwihare, and AmyMarie Accardi. Chemical characterization of high molecular weight dissolved organic matter in fresh and marine waters. *Geochimica et Cosmochimica Acta*, 66(6):955–962, 2002.

[18] Norbert Hertkorn, Ronald Benner, Moritz Frommberger, Philippe Schmitt-Kopplin, Matthias Witt, Karl Kaiser, Antonius Kettrup, and John I. Hedges. Characterization of a major refractory component of marine dissolved organic matter. *Geochimica et Cosmochimica Acta*, 70(12):2990–3010, 2006.

[19] Kalevi Salonen and Taina Hammar. On the importance of dissolved organic matter in the nutrition of zooplankton in some lake waters. *Oceanologica*, 68:246–253, 1986.

[20] John I. Hedges. Chapter 1 Why Dissloved Organics Matter. In Dennis A. Hansell and Craig A. Carlson, editors, *Biogeochemistry of Marine Dissolved Organic Matter*, chapter Why Disslo, pages 1–27. Academic Press, San Diego, CA, 2002.

[21] Rebecca Sutton and Garrison Sposito. Molecular structure in soil humic substances: the new view. *Environmental science & technology*, 39(23):9009–9015, 2005.

[22] IHSS. What are humic substances. http://humic-substances.org/what-are-humic-substances-2/, 2007.

[23] Antonio Nebbioso and Alessandro Piccolo. Basis of a Humeomics Science : Chemical Fractionation and Molecular Characterization of Humic Biosuprastructures. *BioMacromolecules*, 12:1187–1199, 2011.

[24] Xiaodong Ma and Sarah A Green. Chemosphere Fractionation and spectroscopic properties of fulvic acid and its extract. *Chemosphere*, 72:1425–1434, 2008.

[25] Alessandro Piccolo. Conformational Arrangement of Dissolved Humic Substances . Influence of Solution Composition on Association of Humic Molecules. *Environmental Science & Technology*, 33(10):1682–1690, 1999.

[26] Andrea A. Andrew, Rossana Del Vecchio, Yi Zhang, Ajit Subramaniam, and Neil V. Blough. Are Extracted Materials Truly Representative of Original Samples? Impact of C18 Extraction on CDOM Optical and Chemical Properties. *Frontiers in Chemistry*, 4(February):1–12, 2016.

[27] Thorsten Dittmar, Boris Koch, Norbert Hertkorn, and Gerhard Kattner. A simple and efficient method for the solid-phase extraction of dissolved organic matter (SPE-DOM) from seawater. *Limnology and Oceanography: Methods*, 6:230–235, 2008.

[28] John I. Hedges. Global biogeochemical cycles: progress and problems. *Marine Chemistry*, 39(1-3):67–93, 1992.

[29] Melissa C. Kido Soule, Krista Longnecker, Stephen J. Giovannoni, and Elizabeth B. Kujawinski. Impact of instrument and experiment parameters on reproducibility of ultrahigh resolution ESI FT-ICR mass spectra of natural organic matter. *Organic Geochemistry*, 41(8):725–733, 2010.

[30] Neil V. Blough and Rossana Del Vecchio. *Biogeochemistry of Marine Dissolved Organic Matter*. Elsevier, 2002.

[31] Paula G. Coble. Marine optical biogeochemistry: The Chemistry of Ocean Color. *Chemical Reviews*, 107(2):402–418, 2007.

[32] R. G. Zepp, D. J. Erickson III, N. D. Paul, and B. Sulzberger. Interactive effects of solar UV radiation and climate change on biogeochemical cycling. *Photochemical & Photobiological Sciences*, 6(3):286, 2007.

[33] Susan Anderson, Richard Zepp, Jana Machula, Debbie Santavy, Lara Hansen, and Erich Mueller. Indicators of UV Exposure in Corals and Their Relevance to Global Climate Change and Coral Bleaching. *Human and Ecological Risk Assessment: An International Journal*, 7(5):1271–1282, 2001.

[34] Sarah A Green and Neil V Blough. Optical absorption and fluorescence properties of chromophoric dissolved organic matter in natural waters. *Limnology and Oceanography*, 39(8):1903–1916, 1994.

[35] Annick Bricaud, Andre Morel, and Louis Prieur. Absorption by dissolved organic matter of the sea (yellow substance) in the UV and visible domains. *Limnology and Oceanography*, 26(1):43–53, 1981.

[36] Norman Nelson and David Siegel. Chromophoric DOM in the Open Ocean. In Dennis A. Hansell and Craig A. Carlson, editors, *Biogeochemistry of Marine Dissolved Organic Matter*, chapter 11, pages 547–VI. Academic Press, 2002.

[37] Anthony Vodacek, Neil V Bloughl, Michael D Degrandpre, Edward T Peltzer, and Robert K Nelson. Seasonal variation of CDOM and DOC in the Middle Atlantic Bight : Terrestrial inputs and photooxidation. *Limnology and Oceanography*, 42(4):674–686, 1997.

[38] Eurico J D'Sa and Richard L. Miller. Chapter 6 BIO-OPTICAL PROPERTIES OF COASTAL WATERS. In Richard L. Miller, Carlos E. Del Castillo, and Brent A. Mckee, editors, *Remote Sensing of Coastal Aquatic Environments Technologies, Techniques and Applications*, chapter 6, page 134. Springer, Dordrecht, The Netherlands, 7 edition, 2007.

[39] Jason D. Ritchie, David J. Kieber, Kenneth Mopper, Elizabeth C. Minor, John R. Helms, and Aron Stubbins. Absorption spectral slopes and slope ratios as indicators of molecular weight, source, and photobleaching of chromophoric dissolved organic matter. *Limnology and Oceanography*, 53(3):955–969, 2008.

[40] Charles M Sharpless and Neil V Blough. The importance of charge-transfer interactions in determining chromophoric dissolved organic matter (CDOM) optical and photochemical properties. *Environ. Sci.: Processes Impacts*, 16:654–671, 2014.

[41] Rossana Del Vecchio and Neil V. Blough. On the origin of the optical properties of humic substances. *Environmental Science and Technology*, 38(14):3885–3891, 2004.

[42] Paula G Coble. Characterization of marine and terrestrial DOM in seawater using excitation-emission matrix spectroscopy. *Marine Chemistry*, 51:325–346, 1996.

[43] Piotr Kowalczuk, Joanna Ston, William J Cooper, Robert F Whitehead, and Michael J Durako. Characterization of chromophoric dissolved organic matter (CDOM) in the Baltic Sea by excitation emission matrix fluorescence spectroscopy. *Marine Chemistry*, 96:273–292, 2005.

[44] Erin S. Boyle, Nicolas Guerriero, Anthony Thiallet, Rossana Del Vecchio, and Neil V. Blough. Optical Properties of Humic Substances and CDOM: Relation to Structure. *Environmental science & technology*, 43(7):2262–2268, 2009.

[45] Kimmo Hautala, Juhani Peuravuori, and Kalevi Pihlaja. MEASUREMENT OF AQUATIC HUMUS CONTENT BY SPECTROSCOPIC ANALYSES. *Water Research*, 34(1):246–258, 2000.

[46] Yu-ping Chin, George R Aiken, and Edward Loughlin. Molecular Weight , Polydispersity , and Spectroscopic Properties of Aquatic Humic Substances. *Environmental Science & Technology*, 28:1853–1858, 1994.

[47] J Peuravuori and K Pihlaja. Molecular size distribution and spectroscopic properties of aquatic humic substances. *Analytica Chimica Acta*, 337(2):133–149, 1997.

[48] Paula G Coble, Carlos E Del Castillo, and Bernard Avril. Distribution and optical properties of CDOM in the Arabian Sea during the 1995 Southwest Monsoon. *Deep-Sea Research Part II*, 45:2195–2223, 1998.

[49] Juliana D'Andrilli, Christine M. Foreman, Alan G. Marshall, and Diane M. McKnight. Characterization of IHSS pony lake fulvic acid dissolved organic matter by electrospray ionization fourier transform ion cyclotron resonance mass spectrometry and fluorescence spectroscopy. *Organic Geochemistry*, 65:19–28, 2013.

[50] Carmen M Cartisano, Rossana Del Vecchio, Marla R Bianca, and Neil V Blough. Investigating the sources and structure of chromophoric dissolved organic matter (CDOM) in the North Pacific Ocean (NPO) utilizing optical spectroscopy combined with solid phase extraction and borohydride reduction. *Marine Chemistry*, 204:20–35, 2018.

[51] Hussain A. N. Abdulla, Elizabeth C. Minor, and Patrick G. Hatcher. Using two-dimensional correlations of 13C NMR and FTIR to investigate changes in the chemical composition of dissolved organic matter along an estuarine transect. *Environmental Science & Technology*, 44(21):8044–8049, 2010.

[52] N. Hertkorn, M. Harir, B. P. Koch, B. Michalke, and P. Schmitt-Kopplin. High-field NMR spectroscopy and FTICR mass spectrometry: Powerful discovery tools for the molecular level characterization of marine dissolved organic matter. *Biogeosciences*, 10(3):1583–1624, 2013.

[53] Noemie Janot, Pascal E. Reiller, and Gregory V. Korshin. Using Spectrophotometric Titrations To Characterize Humic Acid Reactivity at Environmental Concentrations. *Environmental Science & Technology*, 44(17):6782–6788, 2010.

[54] Thorsten Reemtsma, Anja These, Andreas Springer, and Michael Linscheid. Differences in the molecular composition of fulvic acid size fractions detected by size-exclusion chromatography-on line Fourier transform ion cyclotron resonance (FTICR-) mass spectrometry. *Water Research*, 42(1-2):63–72, 2008.

[55] J.a. Leenheer, M.a. Wilson, and R.L. Malcolm. Presence and potential significance of aromatic-ketone groups in aquatic humic substances. *Organic Geochemistry*, 11(4):273–280, 1987.

[56] DL Macalady and K Walton-Day. Redox Chemistry and Natural Organic Matter (NOM): Geochemists' Dream, Analytical Chemists' Nightmare. In: Tratnyek PG, Grundi TiJ, Haderlein SB (eds). In *Aquatic Redox Chemistry*, chapter Redox Chem, pages 85–111. ACS Symposium Series 1071; American Chemical Society, Washington, D. C, 2011.

[57] Tara Marie Schendorf, Rossana Del Vecchio, Marla Bianca, and Neil V Blough. Combined Effects of pH and Borohydride Reduction on Optical Properties of Humic Substances (HS): A Comparison of Optical Models. *Environmental Science & Technology*, 53(11):6310–6319, 2019.

[58] Norman M Scully, William J Cooper, and Lars J Tranvik. Photochemical effects on microbial activity in natural waters: the interaction of reactive oxygen species and dissolved organic matter. *FEMS Microbiology Ecology*, 46:353–357, 2003.

[59] J A Tossell. Quinone – hydroquinone complexes as model components of humic acids : Theoretical studies of their structure , stability and Visible – UV spectra. *Geochimica et Cosmochimica Acta*, 73(7):2023–2033, 2009.

[60] Jiahai Ma, Rossana Del Vecchio, Kelli S. Golanoski, Erin S. Boyle, and Neil V. Blough. Optical properties of humic substances and CDOM: Effects of borohydride reduction. *Environmental Science and Technology*, 44(14):5395–5402, 2010.

[61] Sabrina M. Phillips and Geoffrey D. Smith. Light Absorption by Charge Transfer Complexes in Brown Carbon Aerosols. *Environmental Science and Technology Letters*, 1(10):382–386, 2014.

[62] Yi Zhang, Rossana Del Vecchio, and Neil V. Blough. Investigating the mechanism of hydrogen peroxide photoproduction by humic substances. *Environmental Science and Technology*, 46(21):11836–11843, 2012.

[63] Michael Aeschbacher, Cornelia Graf, René P Schwarzenbach, and Michael Sander. Antioxidant properties of humic substances. *Environmental Science & Technology*, 46(9):4916–25, 2012.

[64] Daniel R. Baluha, Neil V. Blough, and Rossana Del Vecchio. Selective mass labeling for linking the optical properties of chromophoric dissolved organic matter to structure and composition via ultrahigh resolution electrospray ionization mass spectrometry. *Environmental Science and Technology*, 47(17):9891–9897, 2013.

[65] Marla R Bianca, Daniel R Baluha, Michael Gonsior, Philippe Schmitt-kopplin, Rossana Del Vecchio, and Neil V Blough. Contribution of ketone / aldehyde-containing compounds to the composition and optical properties of Suwannee River fulvic acid revealed by ultrahigh resolution mass spectrometry and deuterium labeling. *Analytical and Bioanalytical Chemistry*, 412:1441–1451, 2020.

[66] Oliver C Zafiriou, Jacques Joussot-Dubien, Richard G Zepp, and G Z Rod. Photochemistry of natural waters. *Environmental Science Technology*, 18(12):358A–371A, 1984.

[67] Jacques Buffle, France L Greter, and Werner Haerdi. Measurement of Complexation Properties of Humic and Fulvic Acids in Natural Waters with Lead and Copper Ion-Selective Electrodes. *Analytical Chemistry*, 49(2):216–222, 1977.

[68] Elizabeth C. Minor, Michael M. Swenson, Bruce M. Mattson, and Alan R. Oyler. Structural characterization of dissolved organic matter: a review of current techniques for isolation and analysis. *Environ. Sci.: Processes Impacts*, 16(9):2064–2079, 2014.

[69] Kenneth Mopper, Aron Stubbins, Jason D. Ritchie, Heidi M. Bialk, and Patrick G. Hatcher. Advanced instrumental approaches for characterization of marine dissolved organic matter: Extraction techniques, mass spectrometry, and nuclear magnetic resonance spectroscopy. *Chemical Reviews*, 107(2):419–442, 2007.

[70] Hussain a N Abdulla, Elizabeth C. Minor, Robert F. Dias, and Patrick G. Hatcher. Changes in the compound classes of dissolved organic matter along an estuarine transect: A study using FTIR and 13C NMR. *Geochimica et Cosmochimica Acta*, 74(13):3815–3838, 2010.

[71] A Piccolo and M Spiteller. Electrospray ionization mass spectrometry of terrestrial humic substances and their size fractions. *Analytical and Bioanalytical Chemistry*, 377:1047–1059, 2003.

[72] Boris P. Koch, Matthias Witt, Ralph Engbrodt, Thorsten Dittmar, and Gerhard Kattner. Molecular formulae of marine and terrigenous dissolved organic matter detected by electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *Geochimica et Cosmochimica Acta*, 69(13):3299–3308, 2005.

[73] Colleen E. Rostad and Jerry a. Leenheer. Factors that affect molecular weight distribution of Suwannee river fulvic acid as determined by electrospray ionization/mass spectrometry. *Analytica Chimica Acta*, 523(2):269–278, 2004.

[74] Jerry A Leenheer, Colleen E Rostad, Paul M Gates, Edward T Furlong, and Imma Ferrer. Molecular Resolution and Fragmentation of Fulvic Acid by Electrospray Ionization / Multistage Tandem Mass Spectrometry. *Anal. Chem.*, 73:1461–1471, 2001.

[75] Nadja B Cech and Christie G Enke. PRACTICAL IMPLICATIONS OF SOME RECENT STUDIES IN ELECTROSPRAY IONIZATION FUNDAMENTALS. *Mass spectrometry reviews*, 20:362–387, 2001.

[76] Peter Nemes, Ioan Marginean, and Akos Vertes. Spraying Mode Effect on Droplet Formation and Ion Chemistry in Electrosprays. *Anal. Chem.*, 79:3105–3116, 2007.

[77] Edmond De Hoffmann and Vincent Stroobant. *Mass Spectrometry - Principles and Applications.* John Wiley & Sons Ltd, West Sussex, England, 3 edition, 2007.

[78] Risto Kostiainen and Tiina J Kauppila. Effect of eluent on the ionization process in liquid chromatography – mass spectrometry. *Journal of Chromatography A*, 1216:685–699, 2009.

[79] William C Hockaday, Jeremiah M Purcell, Alan G Marshall, Jeffery A Baldock, and Patrick G Hatcher. Electrospray and photoionization mass spectrometry for the characterization of organic matter in natural waters: a qualitative assessment. *Limnology and Oceanography*, 7(1):81–95, 2009.

[80] Alexandra C. Stenson, William M. Landing, Alan G. Marshall, and William T. Cooper. Ionization and fragmentation of humic substances in electrospray ionization Fourier transform-ion cyclotron resonance mass spectrometry. *Analytical Chemistry*, 74(17):4397–4409, 2002.

[81] D Cao, H Huang, M Hu, L Cui, F Geng, Z Rao, H Niu, Y Cai, and Y Kang. Comprehensive characterization of natural organic matter by MALDI- and ESI-Fourier transform ion cyclotron resonance mass spectrometry. *Anal. Chim. Acta*, 866:48–58, 2015.

[82] Michael Gonsior, Matthew Zwartjes, William J. Cooper, Weihua Song, Kenneth P. Ishida, Linda Y. Tseng, Matthew K. Jeung, Diego Rosso, Norbert Hertkorn, and Philippe Schmitt-Kopplin. Molecular characterization of effluent organic matter identified by ultrahigh resolution mass spectrometry. *Water Research*, 45(9):2943–2953, 2011.

[83] Sepehr Shakeri Yekta, Michael Gonsior, Philippe Schmitt-Kopplin, and Bo H. Svensson. Characterization of dissolved organic matter in full scale continuous stirred tank biogas reactors using ultrahigh resolution mass spectrometry: A qualitative overview. *Environmental Science and Technology*, 46(22):12711–12719, 2012.

[84] Konstantin O Nagornov, Anton N Kozhinov, and Yury O Tsybin. Fourier Transform Ion Cyclotron Resonance Mass Spectrometry at the Cyclotron Frequency. *American Society for Mass Spectrometry*, 28:768–780, 2017.

[85] Michael P. Balogh. Debating Resolution and Mass Accuracy. *LCGC North America*, 22(2), 2004.

[86] I Jonathan Amster. Fourier Transform Mass Spectrometry. *Journal of Mass Spectrometry*, 31:1325–1337, 1996.

[87] Alan G Marshall, Christopher L Hendrickson, and George S Jackson. FOURIER TRANSFORM ION CYCLOTRON RESONANCE MASS SPECTROMETRY : A PRIMER. *Mass Spectrometry Reviews*, 17:1–35, 1998.

[88] E M Thurman, R.L. Wershaw, R L Malcolm, and D J Pinckney. Molecular size of aquatic humic substances. *Organic Geochemistry*, 4:27–35, 1982.

[89] Elizabeth B Kujawinski, Patrick G Hatcher, and Michael A Freitas. High-Resolution Fourier Transform Ion Cyclotron Resonance Mass Spectrometry of Humic and Fulvic Acids : Improvements and Comparisons as humic acid extracts has been elusive because of. *Anal. Chem.*, 74:413–419, 2002.

[90] Andrew D. Southam, Tristan G. Payne, Helen J. Cooper, Theodoros N. Arvanitis, and Mark R. Viant. Dynamic range and mass accuracy of wide-scan direct infusion nanoelectrospray fourier transform ion cyclotron resonance mass spectrometry-based metabolomics increased by the spectral stitching method. *Analytical Chemistry*, 79(12):4595–4602, 2007.

[91] Boris P Koch, Thorsten Dittmar, Matthias Witt, and Gerhard Kattner. Fundamentals of Molecular Formula Assignment to Ultrahigh Resolution Mass Data of Natural Organic Matter. *Anal. Chem.*, 79(4):1758–1763, 2007.

[92] Elizabeth B. Kujawinski, Krista Longnecker, Neil V. Blough, Rossana Del Vecchio, Liam Finlay, Joshua B. Kitner, and Stephen J. Giovannoni. Identification of possible source markers in marine dissolved organic matter using ultrahigh resolution mass spectrometry. *Geochimica et Cosmochimica Acta*, 73(15):4384–4399, 2009.

[93] Peter Herzsprung, Norbert Hertkorn, Wolf von Tümpling, Mourad Harir, Kurt Friese, and Philippe Schmitt-Kopplin. Understanding molecular formula assignment of Fourier transform ion cyclotron resonance mass spectrometry data of natural organic matter from a chemical point of view. *Analytical and Bioanalytical Chemistry*, 406(30):7977–7987, 2014.

[94] Elizabeth B Kujawinski and Mark D Behn. Automated Analysis of Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectra of Natural Organic Matter. *Anal. Chem.*, 78:4363–4373, 2006.

[95] Rachel L Sleighter and Patrick G Hatcher. SPECIAL FEATURE : The application of electrospray ionization coupled to ultrahigh resolution mass spectrometry for the molecular characterization of natural organic matter. *Journal of Mass Spectrometry*, 42:559–574, 2007.

[96] SM Mugo and CS Bottaro. Characterization of humic substances by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry*, 18:2375–2382, 2004.

[97] E. Michael Perdue and Nelson W Green. Isobaric Molecular Formulae of C, H, and O – A View from the Negative Quadrants of van Krevelen Space. *Anal. Chem.*, 87(10):5079–5085, 2015.

[98] Nelson W. Green and E. Michael Perdue. Fast Graphically Inspired Algorithm for Assignment of Molecular Formulae in Ultrahigh Resolution Mass Spectrometry. *Anal. Chem.*, 87(10):5086–5094, 2015.

[99] Peter Herzsprung, Norbert Hertkorn, Wolf Von Tümpling, Mourad Harir, Kurt Friese, and Philippe Schmitt-kopplin. Molecular formula assignment for dissolved organic matter (DOM) using high-field FT-ICR-MS: chemical perspective and validation of sulphur-rich organic components (CHOS) in pit lake samples. *Anal Bioanal Chem*, 408(10):2461–2469, 2016.

[100] Anja These and Thorsten Reemtsma. Limitations of Electrospray Ionization of Fulvic and Humic Acids as Visible from Size Exclusion Chromatography with Organic Carbon and Mass Spectrometric Detection. *Anal. Chem.*, 75(22):6275–6281, 2003.

[101] Tara Marie Schendorf, Rossana Del Vecchio, Kevin Koech, and Neil V Blough. A standard protocol for NaBH 4 reduction of CDOM and HS. *Limnology and Oceanography: Methods*, 14(6):414–423, 2016.

[102] Donald P Morris and Bruce R Hargreaves. The role of photochemical degradation of dissolved organic carbon in regulating the UV transparency of three lakes on the Pocono Plateau. *Limnology and Oceanography*, 42(2):239–249, 1997.

[103] Huixiang Xie, Oliver C Zafiriou, Wei-Jun Cai, and Yongchen Wang. Photooxidation and Its Effects on the Carboxyl Content of Dissolved Organic Matter in Two Coastal Rivers in the Southeastern United. *Environmental Science and Technology*, 38(15):4113–4119, 2004.

[104] Stephanie M Berg, Quinn T Whiting, Joseph A Herrli, Ronan Winkels, Kristine H Wammer, and Christina K Remucal. The Role of Dissolved Organic Matter Composition in Determining Photochemical Reactivity at the Molecular Level. *Environmental Science & Technology*, 53:11725–11734, 2019.

[105] Philippe Schmitt-Kopplin, Norbert Hertkorn, Hans-Rolf Schulten, and Antonius Kettrup. Structural Changes in a Dissolved Soil Humic Acid during Photochemical Degradation Processes under O2 and N2 Atmosphere. *Environmental Science and Technology*, 32:2531–2541, 1998.

[106] Derek C Waggoner, Andrew S Wozniak, Rose M Cory, and Patrick G Hatcher. The role of reactive oxygen species in the degradation of lignin derived dissolved organic matter. *Geochimica et Cosmochimica Acta*, 208:171–184, 2017.

[107] Aka Marcel Kouassi, Rod G Zika, and John M C Plane. Light-induced alteration of the photophysical properties of dissolved organic matter in seawater : Part II . Estimates of the environmental rates of the natural water fluorescence. *Netherlands Journal of Sea Research*, 27(1):33–41, 1990.

[108] S Twardowski and Percy L Donaghay. Separating in situ and terrigenous sources of absorption by dissolved materials in coastal waters. *JOURNAL OF GEOPHYSICAL RESEARCH*, 106(C2):2545–2560, 2001.

[109] Michael Gonsior, Barrie M Peake, William T Cooper, David Podgorski, Juliana D'Andrilli, and William J Cooper. Photochemically induced changes in dissolved organic matter identified by ultrahigh resolution fourier transform ion cyclotron resonance mass spectrometry. *Environmental science & technology*, 43:698–703, 2009.

[110] Rossana Del Vecchio and Neil V Blough. Photobleaching of chromophoric dissolved organic matter in natural waters : kinetics and modeling. *Marine Chemistry*, 78:231–253, 2002.

[111] David J Kieber, Julie Mcdaniel, and Kenneth Mopper. Photochemical source of biological substrates in sea water : implications. *Nature*, 341(19):637–639, 1989.

[112] Robert J Kieber, Xianliang Zhou, and Kenneth Mopper. Formation of carbonyl compounds from UV-induced photodegradation of humic substances in natural waters: Fate of riverine carbon in the sea. *Limnology and Oceanography*, 35(7):1503–1515, 1990.

[113] Lars Tranvik and Stefan Kokalj. Decreased biodegradability of algal DOC due to interactive effects of UV radiation and humic matter. *Aquatic Microbial Ecology*, 14:301–307, 1998.

[114] Hongmei Chen, Hussain A N Abdulla, Rebecca L Sanders, Satish C B Myneni, Kenneth Mopper, and Patrick G Hatcher. Production of Black Carbon-like and Aliphatic Molecules from Terrestrial Dissolved Organic Matter in the Presence of Sunlight and Iron. *Environmental Science and Technology Letters*, 1:399–404, 2014.

[115] Kenneth Mopper, Xianliang Zhou, Robert Kieber, David Kieber, Richard Sikorski, and Ronald Jones. Photochemical degradation of dissolved organic carbon and its impact on the ocean carbon cycle. *Nature*, 353(5):60–62, 1991.

[116] Elizabeth B. Kujawinski, Rossana Del Vecchio, Neil V. Blough, Geoffrey C. Klein, and Alan G. Marshall. Probing molecular-level transformations of dissolved organic matter: Insights on photochemical degradation and protozoan modification of DOM from electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *Marine Chemistry*, 92(1-4 SPEC. ISS.):23–37, 2004.

[117] Charles M Sharpless, Michael Aeschbacher, Sarah E Page, Jannis Wenk, Michael Sander, and Kristopher Mcneill. Photooxidation-Induced Changes in Optical, Electrochemical, and Photochemical Properties of Humic Substances. *Environmental Science and Technology*, 48:2688–2696, 2014.

[118] Luc Rainville, Steven R. Jayne, and Meghan F. Cronin. Variations of the North Pacific Subtropical Mode Water from Direct Observations. *Journal of Climate*, 27(8):2842–2860, 2014.

[119] Hongyu Li and Elizabeth C Minor. Environmental Science Processes & Impacts Dissolved organic matter in Lake Superior : insights into the e ff ects of extraction methods on chemical composition. *Environ. Sci.: Processes Impacts*, 17:1829–1840, 2015.

[120] U Sarmiento and J L Siegenthaler. Atmospheric carbon dioxide and the ocean. *Nature*, 365(September):119–125, 1993.

[121] Norman B. Nelson and David A. Siegel. Chapter 11 - Chromophoric DOM in the Open Ocean. In *Biogeochemistry of Marine Dissolved Organic Matter*, pages 547–554;555–578. 2002.

[122] Xiaoyan Cao, George R Aiken, Kenna D Butler, Thomas G Huntington, William M Balch, Jingdong Mao, and Klaus Schmidt-Rohr. Evidence for major input of riverine organic matter into the ocean. *Organic Geochemistry*, 116:62–76, 2018.

[123] Oliver J Lechtenfeld, Gerhard Kattner, Ruth Flerus, S Leigh Mccallister, Philippe Schmitt-Kopplin, and Boris P Koch. Molecular transformation and degradation of refractory dissolved organic matter in the Atlantic and Southern Ocean. *Geochimica et Cosmochimica Acta*, 126:321–337, 2014.

[124] Jeffrey A Hawkes, Claudia Patriarca, Per J R Sj, Lars J Tranvik, and Jonas Bergquist. Extreme isomeric complexity of dissolved organic matter found across aquatic environments. *Limnology and Oceanography Letters*, 3:21–30, 2018.

[125] Norbert Hertkorn, Moritz Frommberger, Matthias Witt, Boris P. Koch, Philippe Schmitt-Kopplin, and E. Michael Perdue. Natural Organic Matter and the Event Horizon of Mass Spectrometry. *Anal. Chem.*, 80(23):8908–8919, 2008.

[126] Taylor A Brown, Benjamin A Jackson, Benjamin J Bythell, and Alexandra C Stenson. Benefits of multidimensional fractionation for the study and characterization of natural organic matter. *Journal of Chromatography A*, 1470:84–96, 2016.

[127] Richard G. Zepp, Wade M. Sheldon, and Mary Ann Moran. Dissolved organic fluorophores in southeastern US coastal waters: correction method for eliminating Rayleigh and Raman scattering peaks in excitation–emission matrices. *Marine Chemistry*, 89(1-4):15–36, 2004.

[128] Patrick Louchouarn, Steve Opsahl, and Ronald Benner. Isolation and Quantification of Dissolved Lignin from Natural Waters Using Solid-Phase Extraction and GC / MS. *Analytical Chemistry*, 72(13):2780–2787, 2000.

[129] Peter J. Hernes and Ronald Benner. Transport and diagenesis of dissolved and particulate terrigenous organic matter in the North Pacific Ocean. *Deep-Sea Research Part I: Oceanographic Research Papers*, 49(12):2119–2132, 2002.